



Comparing connected structures in ensemble of random fields

Guillaume Rongier, Pauline Collon, Philippe Renard, Julien Straubhaar, Judith Sausse

► To cite this version:

Guillaume Rongier, Pauline Collon, Philippe Renard, Julien Straubhaar, Judith Sausse. Comparing connected structures in ensemble of random fields. *Advances in Water Resources*, 2016, 96, pp.145-169. 10.1016/j.advwatres.2016.07.008 . hal-01351694

HAL Id: hal-01351694

<https://hal.univ-lorraine.fr/hal-01351694>

Submitted on 5 Aug 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparing connected structures in ensemble of random fields

Guillaume Rongier^{1,2}, Pauline Collon¹, Philippe Renard², Julien Straubhaar², and Judith Sausse¹

¹GeoRessources (UMR 7359, Université de Lorraine / CNRS / CREGU), Vandoeuvre-lès-Nancy, F-54518 France

²Centre d'Hydrogéologie et de Géothermie, Université de Neuchâtel, 11 rue Emile-Argand, 2000 Neuchâtel, Switzerland

Abstract Very different connectivity patterns may arise from using different simulation methods or sets of parameters, and therefore different flow properties. This paper proposes a systematic method to compare ensemble of categorical simulations from a static connectivity point of view. The differences of static connectivity cannot always be distinguished using two point statistics. In addition, multiple-point histograms only provide a statistical comparison of patterns regardless of the connectivity. Thus, we propose to characterize the static connectivity from a set of 12 indicators based on the connected components of the realizations. Some indicators describe the spatial repartition of the connected components, others their global shape or their topology through the component skeletons. We also gather all the indicators into dissimilarity values to easily compare hundreds of realizations. Heat maps and multidimensional scaling then facilitate the dissimilarity analysis. The application to a synthetic case highlights the impact of the grid size on the connectivity and the indicators. Such impact disappears when comparing samples of the realizations with the same sizes. The method is then able to rank realizations from a referring model based on their static connectivity. This application also gives rise to more practical advices. The multidimensional scaling appears as a powerful visualization tool, but it also induces dissimilarity misrepresentations: it should always be interpreted cautiously with a look at the point position confidence. The heat map displays the real dissimilarities and is more appropriate for a detailed analysis. The comparison with a multiple-point histogram method shows the benefit of the connected components: the large-scale connectivity seems better characterized by our indicators, especially the skeleton indicators.

Keywords

Stochastic simulations
Comparison
Static connectivity
Indicators
Dissimilarity

INTRODUCTION

Connectivity is a key aspect of a geological study for its influence on fluid circulations. From a reservoir engineering perspective, it relates to geological structures with high and low permeabilities. But it also relates to the spatial distribution of these structures and the resulting inter-connections, which define the static connectivity. An incorrect connection can bias the results of the flow simulations [Journal and Alabert, 1990, Gómez-Hernández and Wen, 1998]. Reproducing the geological bodies together with their relations is so of prime importance [e.g., Deutsch and Hewett, 1996, King and Mark, 1999].

Stochastic simulations aim at generating possible representations of the geological bodies with respect to the available data. Several methods exist, with a usual separation in two main categories:

- Pixel-based methods simulate one cell at a time, based on a prior model describing the structures of interest. In sequential indicator simulation (SIS) [Deutsch and Journel, 1992], the prior is a variogram built upon the two-point statistics of the data. Hard data conditioning with such method is easy. But the simulated structures do not look like geological bodies. This is especially true for bodies with curvilinear geometries such as channels, whose continuity is badly preserved. The plurigaussian simulation (PGS) [Galli et al., 1994] limits

this difficulty by accounting for the facies relationships. Multiple-point simulations (MPS) go a step further by borrowing multiple-point statistics not from the data but from an external representation of the expected geology, the training image (TI) [Guardiano and Srivastava, 1993].

- Object-based methods rely on the definition of geometric forms and their associated parameters. Each form represents a particular geological body [e.g., Viseur, 2001, Deutsch and Tran, 2002]. The objects are then randomly placed in the domain of interest with parameters drawn in statistical laws. More recent approaches introduce some genetic aspects to improve the object organization [e.g., Lopez, 2003, Pyrcz et al., 2009]. They provide more geologically consistent results. For instance channel continuity and relationships are better preserved than with pixel-based methods. But this is at the cost of the ease of parametrization. And object-based approaches have difficulty to condition the objects to data.

All these methods have advantages and drawbacks. This will influence the choice of a method and its parameter values when dealing with a case study.

But few work aims at systematically analyzing the quality of a set of realizations regarding their static connectivity. The quality control often consists in comparing the histogram and variogram of several realizations with those of the data, or

of the training image if any [e.g., [Strebelle, 2002](#), [Mariethoz et al., 2010](#), [Tahmasebi et al., 2012](#)]. If more than the first two-order statistics are necessary to simulate geological bodies [e.g., [Guardiano and Srivastava, 1993](#), [Journel, 2004](#)], the same conclusion must apply when comparing realizations. Some authors propose to also use the higher-order statistics for quality analysis. [Boisvert et al. \[2010\]](#) and [Tan et al. \[2014\]](#) propose to analyze the multiple-point histogram. [De Iaco and Maggio \[2011\]](#) and [De Iaco \[2013\]](#) also explore the multiple-point statistics with high-order cumulants.

The purpose of most simulation methods is to reproduce statistics from a prior. Analyzing statistics highlights the method success in this reproduction, not if the realizations are geologically consistent. To do that, the statistical analysis is often completed by a visual evaluation of the global structures. The geological structures are compared to what is expected from the known geology, with a focus on the further use of the realizations. This use is often related to fluid circulations, and requires an assessment of the static connectivity, which is not directly imposed by the simulation methods contrary to the statistics. But a visual analysis remains subjective and limited to a few realizations, often in two-dimensions [e.g., [Yin et al., 2009](#), [Tahmasebi et al., 2012](#)].

Yet, some studies focus on analyzing the connectivity of the realization bodies. For instance [Meerschman et al. \[2012\]](#) use the connectivity function with the histogram and variogram to analyze the simulation parameter impact for the Direct Sampling MPS method [[Mariethoz et al., 2010](#)]. [Deutsch \[1998\]](#) uses directly the connected components determined from lithofacies, porosity and permeability models. He computes indicators such as the number of connected components or their sizes to rank the realizations. [De Iaco and Maggio \[2011\]](#) and [De Iaco \[2013\]](#) also use some measures related to the connected components, such as their number or their mean surface and volume. [Comunian et al. \[2012\]](#) rely on some of the previous indicators to analyze the quality of three-dimensional structures simulated from two-dimensional training-images. They also consider the equivalent hydraulic conductivity tensor as an indicator. However, this requires to have an idea of the hydraulic conductivities for the simulated facies.

Connected components enable to characterize the geometry and topology of the geological bodies, which is the purpose of the visual comparison of realizations. They also enable to study the static connectivity of the geological bodies, while being easy to compute. Contrary to a visual analysis of the realization, indicators from connected components are unbiased and can compare many realizations. Contrary to statistical or hydraulic property indicators, they focus on the sedimentary bodies by characterizing their connectivity and are more easy to apprehend. However, current methods based on the connected components are limited to few simple indicators, often analyzed independently.

This leads to the question of the result visualization to analyze more effectively the indicators. [Scheidt and Caers \[2009\]](#) and [Tan et al. \[2014\]](#) both rely on the computation of dissimilarity values between the realizations. Those dissimilarities are computed based on the quality indicators measured on each realization. They are then visualized based on a Multi-Dimensional Scaling (MDS) [e.g., [Torgerson, 1952](#), [Shepard, 1962a,b](#)]. MDS represents the realizations as points, with the distance between the points as close as possible to the dissimilarities. The global analysis of the realization dissimilarities is so easier.

The present work aims at analyzing and discussing a set of indicators to quantify the quality of stochastic simulations from the viewpoint of static connectivity. This method performs on categorical three-dimensional images representing the facies constituting the geological bodies of interest. It can be applied on realizations from one or several stochastic simulation methods and/or parameter values. Conceptual images representing ideally the structures to simulate can also be considered. The chosen set of indicators relies on quantitative measurements on connected components and their skeletons (section 1). The indicators are used in dissimilarity computations to analyze the quality more directly (section 2). Several realizations obtained with different simulation methods (section 3.1) are then used to test the method and compare it to the multiple-point histograms (section 3), and discuss the results (section 4).

1 INDICATORS TO MEASURE SIMULATION QUALITY

The quality analysis fits in a stochastic process implying the simulation of many realizations in a grid. It further investigates the differences of static connectivity between these realizations.

1.1 About grids and grid cells

Many methods to simulate geological structures rely on a discretized representation of the domain of interest: a grid. The grid is a volumetric mesh composed of simple elements, hereinafter called cells.

Many types of grid exist, with different cell types (e.g., tetrahedron or hexahedron). Most of the stochastic simulation methods rely on hexahedral grids, either regular or irregular. Irregular hexahedral grids help to be as conform as possible to the geological structures such as horizons and faults. The sedimentary bodies are then simulated within the parametric space of the grid [e.g., [Shtuka et al., 1996](#)]. The parametric space mimics a deposition space to get rid of the deformation and faulting occurring after deposition and materialized in the grid geometry.

Consequently, the indicators are computed on hexahedral grids, both regular and irregular. Similarly to the simulation, the indicator computation is done in the parametric space of the grid. Thus, the indicators based on volumes or surfaces are rather computed using number of cells and number of faces. This avoids biases related to different grid geometries, which give different indicator values even if the objects are the same when transferred in the same grid.

Within a grid, the cells are connected one to another by their faces, their edges and/or their corners [[Renard and Allard, 2013](#)]. In the case of the hexahedral grids used for this work, one cell has three possible neighborhoods (figure 1):

1. One neighborhood composed of six face-connected cells.
2. One neighborhood composed of eighteen face- and edge-connected cells.
3. One neighborhood composed of twenty-six face-, edge- and corner-connected cells.

This definition of the connectivity between a cell and its neighborhood can be extended to form connected components.

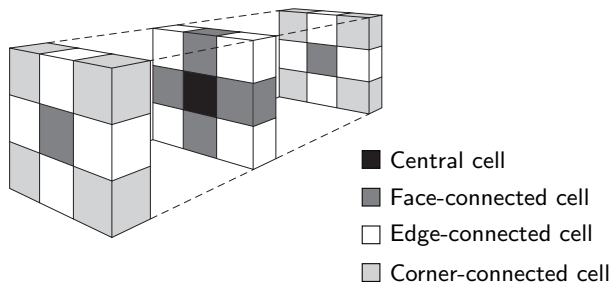


Figure 1 Possible neighborhoods for a given central cell in a regular grid (modified from Deutsch [1998]).

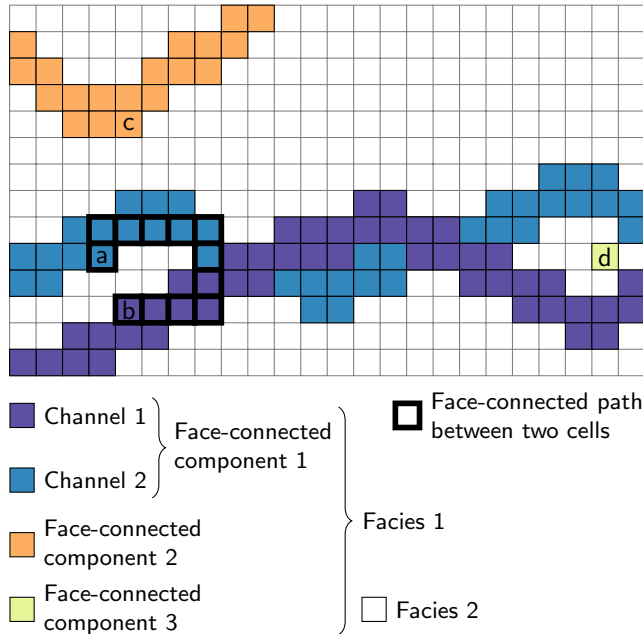


Figure 2 Connected components of a given facies in a two-dimensional structured grid. The cells *a* and *b* are connected and belong to the same connected body. There is no possible connected path between those cells and *c*, which belongs to another connected body. The cell *d* constitutes a third connected body in the case of a face-connected neighborhood. In the case of an edge- or corner-connected neighborhood, *d* belongs to the connected body 1.

1.2 Basic element: the connected component

The connected components result from the widening of the neighborhoods. They rely on the following definition of the connectivity between two cells: two cells belonging to the same facies are connected if a path of neighboring cells remaining within the same facies exists (figure 2). Applying this definition to all the cells of a facies gives the connected components of this facies.

This leads to a distinction between the geological objects, such as a channel or a crevasse splay, and the connected components. Indeed, the geological objects often tend to cross each others, giving one connected body where there is in fact several geological objects (figure 2). The range of possible shapes is larger for the connected components than for the individual object. This aspect complicates the comparison between images. But determining the connected components is far easier than trying to retrieve the geological objects. This is also close to the functioning of pixel-based methods, which do not try to reproduce geological objects but groups of cells, and therefore connected components.

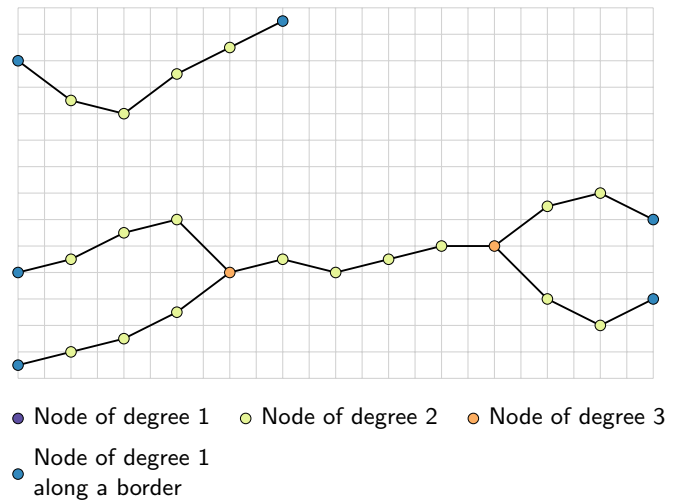


Figure 3 Example of skeletons for the connected components of the figure 2. Here the nodes connected to only one segment – the nodes of degree 1 – are all along a grid border. Two nodes of degree three highlight the local disconnections between the channels at the bottom. The connected component at the top has no node of degree higher than two, which shows the complete connectivity of all its cells, even locally.

1.3 Basic element: the skeleton

A curve-skeleton – simply called here skeleton – is a thin one-dimensional representation of a three-dimensional shape. It is composed of nodes linked together by one or more segments (figure 3). The degree of a node is the number of segments connected to that node. Skeletons are often used to study some geometrical and topological features of a shape. Here the skeletons are those of the connected components. They enable to further characterize the global shape of the connected components, while giving more details about their topology than indicators directly computed on the components.

Several methods exist to compute skeletons [e.g., Serra, 1983, Jain, 1989, Brandt and Algazi, 1992]. The method considered for this work is based on slicing the grid along a given axis. The grid is subdivided into parallel slices of a given thickness. On each slice the connected components are computed and one node is assigned to each component. The nodes are then linked by computing the connected components over two adjacent slices. If two components from two slices form one connected component when the slices are combined, their nodes are linked. If they form several components, their nodes are not linked.

1.4 Indicators

The indicators studied in this paper focus on analyzing the connectivity of the geological bodies within a three-dimensional image. This static connectivity analysis is possible thanks to the connected components. All the indicators are quite simple and each one gives only partial information about the connectivity and its structure. But their combination provides a more detailed characterization.

Appendix A defines in detail all the indicators. Table 1 summarizes the indicator definition, by focusing on their relationship with the connected components. We distinguish three categories of indicators:

Global indicators: The global indicators characterize a facies

Table 1 The set of indicators with their definitions. The indicator definitions is given by a numerator and a denominator because the majority of the indicators comes from a ratio. Some ratios are computed on a single connected component, and their values are combined to obtain one indicator value per facies. We use the term “component” instead of “connected component” for the sake of simplicity.

Category	Indicator	Symbol	Numerator	Denominator	Value for a facies
Global indicators	Facies proportion	p	Number of component cells	Number of cells of the grid	Sum of all the component values
	Facies adjacency proportions	p^a	Number of component cells adjacent to a cell of a given other facies	Number of cells of the facies adjacent to a cell of any other facies	Sum of all the component values
	Facies connection probability	Γ	Squared number of component cells	Squared number of cells of the facies	Sum of all the component values
	Connected component density	ϵ	Number of components	Number of cells of the grid	–
	Unit component proportion	p^u	Number of components of one cell	Number of components	–
Shape indicators	Traversing component proportion	p^c	Number of components linking two opposite borders of the grid	Number of components	–
	Number of component cells	n	Number of component cells	–	Average of the non-unit-component values
	Box ratio	β	Number of cells of a component	Number of cells of the axis-aligned bounding box of the component	Average of the non-unit-component values
	Faces/cells ratio	ζ	Number of faces composing a component surface	Number of cells of a component	Average of the non-unit-component values
	Sphericity	ϕ	Surface area of a sphere	Surface area of the connected component	Average of the non-unit-component values
Skeleton indicators	Inverse branch tortuosity	t	Distance between the extremities of a branch	Branch curvilinear length	Average of all the branch values
	Node degree proportions	p^n	Number of node connected to n segments	Total number of node for all the skeletons	Sum of all the skeleton values

and not necessarily an individual connected component. Among them, the facies proportion is a classical indicator to compare realizations. Some others, such as the facies connection probability [Renard and Allard, 2013], the connected component density or the traversing component proportion give an idea of the global connectivity.

Shape indicators: Global measures such as facies proportions are not sufficient to characterize precisely the impact of the related facies on the flow [e.g., Western et al., 2001, Mariethoz, 2009]. In particular, Oriani and Renard [2014] showed the influence of the connected component geometry – i.e., their shape – on the equivalent hydraulic conductivity, and therefore on the flow behavior. The shape indicators characterize the connected component shape through simple surface and volumetric measures. They all give one value per component. The arithmetic mean of those values provides a value of the indicator for a given facies. This makes the indicator comparison easier.

Skeleton indicators: The skeletons help to better characterize the topology and global geometry of their connected components: their one-dimensional representation is easier to analyze. Here two indicators are introduced. The inverse branch tortuosity characterizes the geometry of the skeleton. Its values for all the branches of all the skeletons related to a facies are averaged to obtain a single value for the facies. It completes the shape indicators in the characterization of the connected component shape. The node degree proportion depicts the topology of the skeletons. It helps to analyze the connectivity more precisely.

2 QUALITY ANALYSIS CONSIDERATIONS

The final purpose of this work is to easily and objectively compare several realizations. The indicators are thus computed on large sets of realizations, which may come from different methods and/or parameters. Then dissimilarity values based on the indicators help to compare the realizations.

2.1 Influence of different grid dimensions

Some cases imply to compare realizations on different grids, and the grids may have different dimensions. For instance in MPS, the training image is often larger than the simulation grid to maximize pattern repeatability.

The grid dimensions influence the size of the traversing connected components, such as channels. This impacts in particular the connected component density and the number of component cells. When the grid size varies along the channel direction, the number of cells for the channels also varies. And even though the number of channels does not necessarily change, the grid volume does, impacting the density. These indicators highlight expected differences in such cases. Their direct use is then detrimental to the quality analysis.

We propose two workarounds to compensate for different grid sizes:

- Either sampling the images from the different grids so that all the samples have the same dimensions. The sample size are the largest dimensions common to all

the grids. Each sample is randomly extracted and each image may be sampled several times to still catch the characteristics of the whole image.

- Or correcting the indicators of the difference between the grid dimensions. The smallest grid dimensions among all the grids form a hypothetical reference grid. The indicators are corrected to their expected value in such reference grid. Appendix B details this correction.

The sampling exempts from correcting the indicators, but it adds a step and requires the analysis of more images, which could slow down the process. If they are valid, the corrections should give similar results than the sampling in a more efficient process.

2.2 Indicator rescaling

The rescaling ensures that the differences between the ranges of indicator values will not affect the comparison. The histogram-based indicators – facies proportion, facies adjacency proportion and node-degree proportion – are not rescaled, to preserve their histogram behavior for the dissimilarity computation (section 2.3). Two methods can be used for rescaling: normalization and standardization.

The normalization method consists in rescaling linearly the indicators values between 0 and 1. The indicator I_i is the i^{th} indicator of the set previously defined. When computed for the facies f of the realization r , we will denote the computed indicator I_{if}^r . The normalization is then obtained by rescaling it between its minimum and maximum values:

$$\text{norm}(I_{if}^r) = \frac{I_{if}^r - m_{if}}{M_{if} - m_{if}} \quad (1)$$

with M_{if} the maximum value for the same indicator and facies among all the images and m_{if} the minimal value for the same indicator and facies among all the images.

The standardization method consists in using reduced-centered indicator values. For an indicator i the standardized value for a facies f of a realization r is obtained using the following formula:

$$\text{stand}(I_{if}^r) = \frac{I_{if}^r - \mu_{if}}{\sigma_{if}} \quad (2)$$

with μ_{if} the mean for the same indicator and facies among all the images and σ_{if} the standard deviation for the same indicator and facies among all the images. Standardization is an interesting option to focus on the indicator variance. The normalization on the other hand decreases the influence of outliers and gives precise limits to the indicator values.

2.3 Dissimilarity calculation

The principle of comparing two images is to determine how dissimilar these images are. The indicators can be seen as coordinates of the compared images. These indicators are heterogeneous: they are either based on histograms or on continuous values. The computation of a dissimilarity value between two images requires a heterogeneous metric.

Following the example of [Wilson and Martinez \[1997\]](#), two different metrics are combined into a heterogeneous Euclidean/Jensen–Shannon metric. It uses the Jensen–Shannon distance, square root of the Jensen–Shannon divergence [[Rao, 1987](#), [Lin, 1991](#)], for the histogram-based indicators – facies proportion,

facies adjacency proportion and node-degree proportion – and the Euclidean distance for all the other indicators. The distance between two images r and s for a given indicator i of a given facies f is given by:

$$d(I_{if}^r, I_{if}^s) = \begin{cases} d_{JS}(I_{if}^r, I_{if}^s) & \text{if } I_{if}^r \text{ and } I_{if}^s \text{ are histograms} \\ d_E(I_{if}^r, I_{if}^s) & \text{if } I_{if}^r \text{ and } I_{if}^s \text{ are continuous values} \end{cases} \quad (3)$$

with I the indicator values. d_{JS} represents the Jensen–Shannon distance:

$$d_{JS}(H_i^r, H_i^s) = \sqrt{\frac{\frac{1}{2} \sum_{j=1}^n \left[H_{ij}^r \log \left(\frac{H_{ij}^r}{\frac{1}{2}(H_{ij}^r + H_{ij}^s)} \right) + H_{ij}^s \log \left(\frac{H_{ij}^s}{\frac{1}{2}(H_{ij}^r + H_{ij}^s)} \right) \right]}{H_{ij}^r \log \left(\frac{H_{ij}^r}{\frac{1}{2}(H_{ij}^r + H_{ij}^s)} \right) + H_{ij}^s \log \left(\frac{H_{ij}^s}{\frac{1}{2}(H_{ij}^r + H_{ij}^s)} \right)}}} \quad (4)$$

with H_i^r and H_i^s the histograms of the indicator i for respectively the images r and s , n the number of classes for each histogram, H_{ij}^r and H_{ij}^s the proportions for the class j in the corresponding histograms. d_E represents the Euclidean distance used with rescaled indicators:

$$d_E(I_{if}^r, I_{if}^s) = \sqrt{(\text{resc}(I_{if}^r) - \text{resc}(I_{if}^s))^2} \quad (5)$$

with I_{if}^r and I_{if}^s the values of the indicator i for the facies f of respectively the images r and s and resc either norm (formula 1) or stand (formula 2). The final dissimilarity δ between two images r and s given their respective sets of indicators I^r and I^s is:

$$\delta(I^r, I^s, \omega, \nu) = \sqrt{\omega_1 d_{JS}(I_1^r, I_1^s)^2 + \sum_{i=2}^{12} \sum_{f=1}^n \omega_i \nu_f d(I_{if}^r, I_{if}^s)^2} \quad (6)$$

with I_1^r and I_1^s the facies proportion histogram for the two images, I_{if}^r and I_{if}^s all the other indicator values depending on the indicator and the facies and n the number of facies. ω represents the set of weights ω_i that control the impact of each indicator. ν represents the set of weights ν_f that control the impact of each facies. Note that the facies proportion histograms are the only indicators with one result for all the facies. Thus the facies proportions are treated differently from all the other indicators. The Jensen–Shannon distance used in that case is slightly modified:

$$d_{JS}(H_i^r, H_i^s, \nu) = \sqrt{\frac{\frac{1}{2} \sum_{f=1}^n \nu_f \left[H_{if}^r \log \left(\frac{H_{if}^r}{\frac{1}{2}(H_{if}^r + H_{if}^s)} \right) + H_{if}^s \log \left(\frac{H_{if}^s}{\frac{1}{2}(H_{if}^r + H_{if}^s)} \right) \right]}{H_{if}^r \log \left(\frac{H_{if}^r}{\frac{1}{2}(H_{if}^r + H_{if}^s)} \right) + H_{if}^s \log \left(\frac{H_{if}^s}{\frac{1}{2}(H_{if}^r + H_{if}^s)} \right)}}} \quad (7)$$

The dissimilarity values computed by formula 6 between all the images constitute a non-negative symmetric matrix. This matrix has a zero diagonal corresponding to the dissimilarity between an image and itself. The dissimilarity matrix can be directly visualized with a heat map or treated by multidimensional scaling to get a more practical visualization.

2.4 Heat map

The heat map is a simple graphical representation of a matrix where the matrix values correspond to colors. In our case, the heat map is a two-dimensional representation. This colored representation highlights patterns in the dissimilarity matrix, either between realizations or between simulation methods. The main advantage of the heat map is to show the real dissimilarity values, contrary to the multidimensional scaling described in the next subsection.

The heat map also enables to classify the images and/or to apply clustering methods on it. A simple yet informative classification is the ranking according to the dissimilarities of the images toward one particular image. When using more advanced clustering methods, the matrix rows and columns are permuted to gather close values into the same cluster.

2.5 Multidimensional scaling

Multidimensional scaling (MDS) [e.g., Torgerson, 1952, 1958, see Cox and Cox, 1994 for a review] is a set of data visualization methods to explore dissimilarities between objects – represented by a dissimilarity matrix – through a dimensionality reduction: it aims at producing a configuration of the objects as optimal as possible in a lower dimensional representation.

2.5.1 Principle and method used

Finding the configuration of the images in a k dimensional representation consists in locating a set of points representing the objects in a k -dimensional Euclidean space – with k being at most equal to the number of images minus one. The point positioning is done so that the Euclidean distance d between two points matches as closely as possible the dissimilarities between the images:

$$d_{r,s} = \sqrt{\sum_{i=1}^k (x_{ri} - x_{si})^2} \quad (8)$$

with r and s two images, k the dimension number of the Euclidean space, x_{ri} and x_{si} the coordinates of respectively r and s in the i -th dimension. The number of dimension k for the MDS representation is an input parameter. When equal to the number of images minus one, the distances are normally equal to the dissimilarities. When k is lower, the MDS misrepresents more or less the dissimilarities.

Several multidimensional scaling methods have been proposed [e.g., Cox and Cox, 1994], depending on the type of dissimilarities and on the way to match the dissimilarities with the distances. The classical scaling [Torgerson, 1952, 1958, Gower, 1966] is the usual method for multidimensional scaling [Scheidt and Caers, 2009, Tan et al., 2014, e.g.,]. It assumes that the dissimilarities already are Euclidean distances. If this assumption can be relaxed to a metric assumption, i.e., the dissimilarities are distances, Euclidean or not, the classical

scaling may further misrepresents dissimilarities based on a heterogeneous metric.

Here we use a different method: the Scaling by MAjorizing a COmplicated Function (SMACOF) [De Leeuw, 1977, De Leeuw and Heiser, 1977, 1980]. Its goal is to get distances as close as possible from the dissimilarities using a majorization, i.e., the optimization of a given objective function called stress, through an iterative process. The stress derives from the squared difference between the dissimilarities and the distances. It is positively defined and equals to 0 only when the distances are equal to the dissimilarities. The optimization process corresponds to a minimization of the stress. The final stress value helps to assess the choice of the number of dimensions: the lower the stress is, the better is that choice.

2.5.2 Validation of the number of dimensions

Following the chosen number of dimensions for the representation, the point configuration matches more or less the dissimilarity values. Verifying that the dimension number is enough for a good match between the dissimilarities and the distances is so of prime importance. Two approaches allow testing the chosen dimension number:

The scree plot: It represents the stress of the SMACOF against the dimension number. The stress follows a globally convex decreasing function that tends toward 0 when the dimension number increases. A stress close or equal to zero means that the higher dimensions are unnecessary to represent the dissimilarities. The best number of dimensions is between the point with the highest flexion of the curve and the beginning of the sill at zero. The dimension value right after the point with the highest flexion is generally enough for a decent representation.

The Shepard diagram: It represents the distances against the dissimilarities. The better the correlation, the better the choice of dimension number.

Two-dimensions are more practical for an analysis purpose. A three-dimensional representation remains a possibility if the improvement is significant enough from a two-dimensional representation.

2.5.3 Estimation of the point position confidence

The point position confidence is another way to assess the MDS ability to represent the dissimilarities. For each point r , an error e highlights the mismatch between the dissimilarities δ and the distances d with all the other points s :

$$e_r = \sum_s |(a\delta_{r,s} + b) - d_{r,s}| \quad (9)$$

with a and b the linear regression coefficients found on the Shepard diagram. This measure gives a more local representation of the miss-representation than the scree plot or the Shepard diagram.

For visualization purpose, that error is then normalized, giving the confidence c for a given image r :

$$c_r = 1 - \frac{e_r - e_{\min}}{e_{\max} - e_{\min}} \quad (10)$$

with e_{\max} and e_{\min} respectively the greatest and the lowest error values amongst the errors of all the images. This confidence can then be attributed to its corresponding point in the

MDS representation through the point transparency: the less transparent the point is, the best the dissimilarities related to this point with all the other points are represented.

3 EXAMPLE OF METHOD APPLICATION

The method, as described in the previous sections, consists in three steps:

1. Indicator computation.
2. Dissimilarity computation in a matrix.
3. Dissimilarity visualization and analysis, especially with multidimensional scaling.

The first two steps were implemented in a C++ plugin for the SKUA-GOCAD geomodeling software [Paradigm, 2015]. The last step was realized using the software environment for statistical computing R [R Core Team, 2012] with the addition of the R packages SMACOF [De Leeuw and Mair, 2009] and ggplot2 [Wickham, 2009].

3.1 Dataset

The dataset falls within the simulation of a channelized system. It contains several realizations representing the same sedimentary environment simulated with different methods. The analysis aims at highlighting the indicator ability to capture the differences of static connectivity between the realizations, and especially between the realizations from different methods. As it concerns a sole case, it would be inappropriate to draw general conclusions on the simulation methods themselves.

The channelized system is composed of sandy channels with levees into a mudstone environment. A conceptual model, called the training image (TI) (figure 4, image at the top), provides an ideal representation of this system. The case study falls within the scope of a MPS study: several simulation methods are used to reproduce the sedimentary bodies observed in the training image. MPS performs better when the training image is larger than the realizations, to ensure enough pattern repeatability. It involves two grids: the first one for the training image (figure 4, image at the top) and the second one for the realizations (figure 4, images at the bottom).

The training grid contains two sets of images:

TI: One object-based realization simulated using the object-based method of the software Petrel [Schlumberger, 2015] (see appendix C, table C.3, for the simulation parameters).

Analog: 100 object-based realizations simulated with the same method and parameters used to simulate the TI (appendix C, table C.3).

The simulation grid contains four sets of images:

DeeSse: 100 MPS realizations simulated with the DeeSse implementation [Straubhaar, 2011] of the direct sampling method [Mariethoz et al., 2010]. Contrary to more traditional MPS methods, the direct sampling bypasses the conditional probability computation and resamples randomly the training image. It relies on the compatibility – measured with a distance – between the conditioning data and the patterns scanned in the training image. The

resampling step selects the first pattern with a distance lower than a given threshold. The training image is the TI and the set of parameters is given in table C.1) in the appendix.

IMPALA: 100 MPS realizations simulated with the method IMPALA [Straubhaar et al., 2011, 2013]. Contrary to the DeeSse, IMPALA still computes the conditional probabilities during the simulation. To improve the efficiency of this computation, the method stores the training image patterns in a list. The training image is scanned once at the beginning and the list is used instead during the simulation. The training image is the TI and the set of parameters is given in table C.2) in the appendix.

OBS: 100 object-based realizations simulated with the same method and parameters used to simulate the TI (appendix C, table C.3).

SIS: 100 sequential indicator simulation realizations simulated using variograms based on the facies in the TI (appendix C, table C.4).

3.2 Analysis setting

The purpose here is to compare the realizations with the TI. It should lead to retain the method and associated parameters that reproduce at best the static connectivity of the TI for the studied case. The indicators used in this case study (table 2) rely on the face-connected components, because the face-connectivity between cells is the most frequently used [Renard and Allard, 2013]. All the indicators are equally considered ($\omega_i = 1$ for all i in formula 6). This avoids any subjective bias that could arise from favoring a given indicator. The mudstone environment is the resultant of the channels and levees placement. It has so no precise shape by itself and may blur the analysis. It gets a weight of 0 while channels and levees both get each a weight of 1 ($v_{mudstone} = 0$, $v_{channel} = 1$ and $v_{levee} = 1$). Channels and levees are considered equally important to reproduce, but this aspect is related to the case study and could be further discussed. The indicators are normalized to cancel the differences of different indicator ranges. Slices of 17 cells along the grid axis with the same orientation than the channels are used for the skeletonization.

Several samples are also randomly extracted from the grids to evaluate the suitability of correcting the indicators when dealing with different grid sizes. The training grid having $500 \times 500 \times 20$ cells and the simulation grid $100 \times 150 \times 30$ cells, the common largest dimensions for the samples are $100 \times 150 \times 20$. The training grid is almost 10 times larger than the simulation grid. Therefore, 20 samples are extracted from the TI and each analog, whereas 2 samples are extracted from each DeeSse, IMPALA, OBS and SIS realization.

3.3 Visual inspection of the realizations

Looking at the connected components (figure 5) highlights some expectations for the dissimilarity analysis. Two aspects must be analyzed: the reproduction of the sedimentary body shapes and the reproduction of their connectivity, especially concerning the channels. In the studied case, the reproduction of the shape is pretty easy to analyze visually. The SIS realizations do not display any objects similar to channel/levee

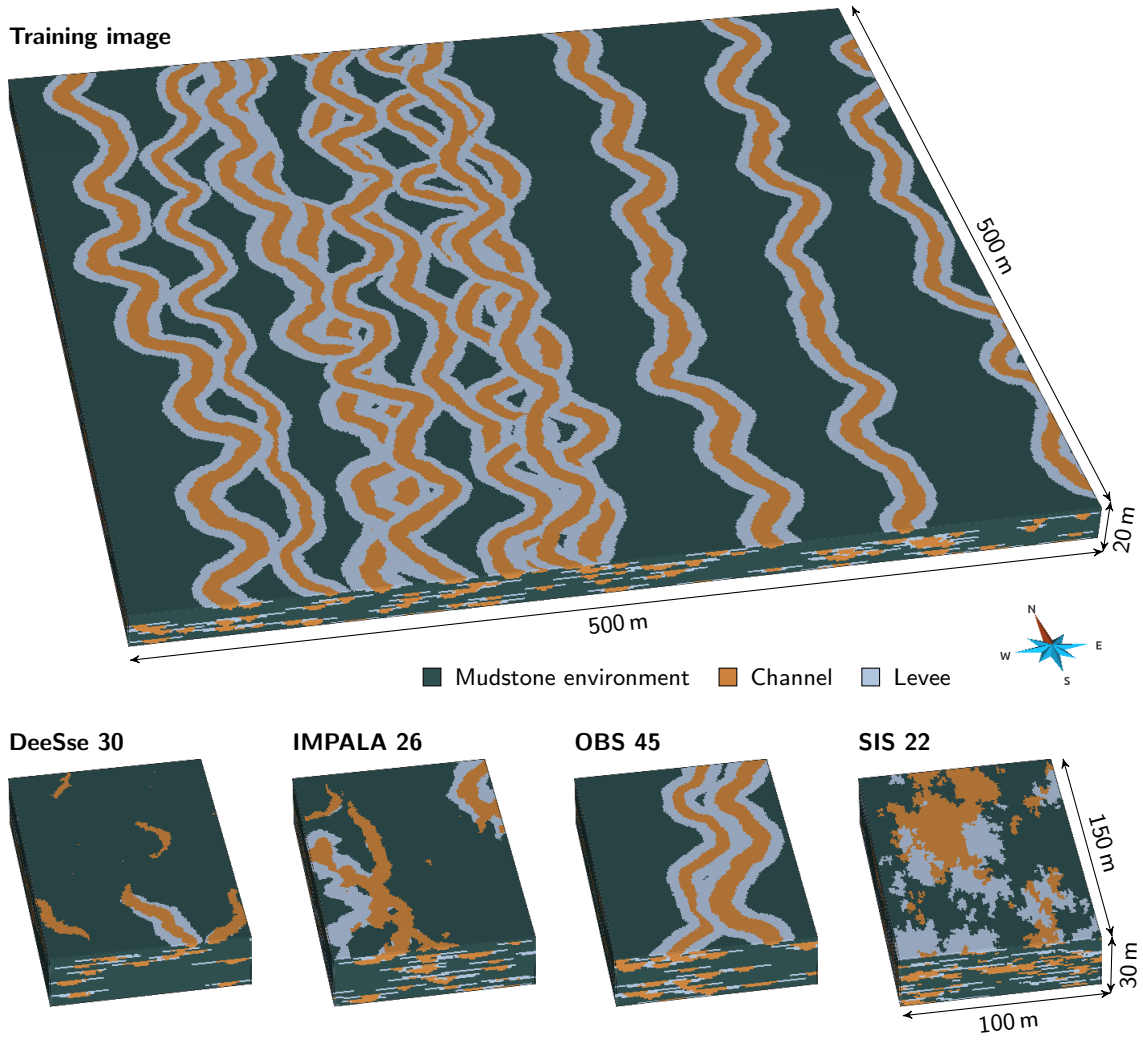


Figure 4 Training image and examples of realizations for each category.

Table 2 Set of indicators used for the case study. The indicator definitions are summarized in table 1 and more detailed descriptions are in appendix A.

Category	Indicator	Symbol	Weight
Global indicators	Facies proportion	p	1
	Facies adjacency proportion	p^a	1
	Facies connection probability	Γ	1
	Connected component density	ϵ	1
	Unit connected component proportion	p^u	1
	Traversing connected component proportion	p^c	1
Shape indicators	Number of connected component cells	n	1
	Box ratio	β	1
	Faces/cells ratio	ζ	1
	Sphericity	ϕ	1
Skeleton indicators	Node degree proportions	p^n	1
	Inverse branch tortuosity	t	1

systems and are so far dissimilar from the TI. The OBS realizations look similar to the TI, which is what is expected considering that they come from the same method and parameters. DeeSse realizations have objects similar to channels, even if some continuity issues appear. They also seem to have an insufficient number of channels. IMPALA realizations have quite

linear objects but which poorly reproduce channel and levee shapes.

Estimating the static connectivity in three-dimensional images is more difficult. The TI channels seem highly connected. The objects in the SIS realizations do not locally intersect like channels do and are far too connected. DeeSse realizations con-

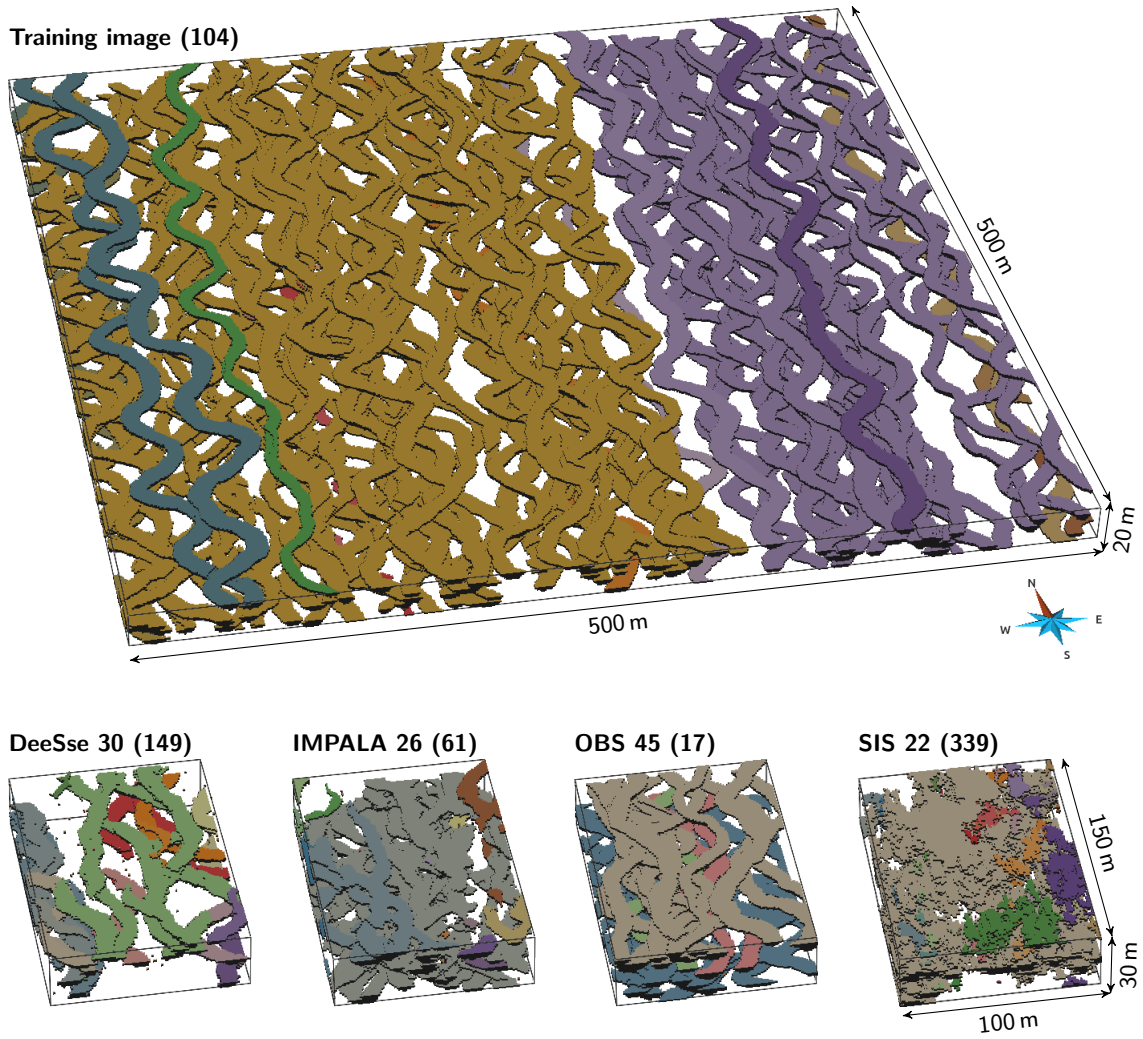


Figure 5 View of all the channel connected components within the TI and examples of realizations for each categories. The number in parentheses are the number of connected components of each image.

tain less objects and seem under-connected compared to the TI. The distinction between OBS and IMPALA realizations is difficult concerning the connectivity. Looking at the skeletons of the connected components (figure 6) corroborates those observations. DeeSse realizations are clearly under-connected compared with the other categories. SIS ones are over-connected. IMPALA realizations seem a bit more connected than OBS ones. The static connectivity within the training image is clearly heterogeneous.

3.4 Effect of different grid dimensions on the analysis

The TI, analogs and OBS realizations come from the same method with the same parameter values. The grid size is the only difference between all these images: the grid of the TI and analogs – the training grid – is larger than that of the OBS realizations – the simulation grid.

This difference of grid dimensions directly impacts the connected component density and the number of connected component cells, which are corrected to take into account such difference (appendix B). But the realizations coming from the same method still differ when looking at the dissimilarities (figure 7, MDS representation from the original images). The OBS realizations within the simulation grid stand out from the TI and the analogs within the training grid. Such difference is

absent from the samples, where all the images have the same size (figure 7, MDS representation from the image samples). The grid size seems to clearly impact the dissimilarity values.

However, both MDS representations (figure 7) have high stress values with two dimensions and can not be fully trusted. The heat maps (figure 8) clarify that situation. The heat map from the original images (figure 8, bottom left) appear non-homogeneous. A red square symbolizes the significant dissimilarities between the TI and analogs on one side and the OBS realizations on the other side. The heat map from the samples is far more homogeneous, without red square. They confirm the impact of the grid size observable on the MDS representations.

Thus, correcting the connected component density and the number of connected component cells is not adequate, and other indicators are impacted by the grid dimensions. The TI, the analogs and the OBS realizations have all similar channel and levee proportions (figure 9). The channels and levees occupy the same volume inside the two grids. But the facies connection probabilities for both channels and levees differ between the realizations in the two grids (figure 9). The probability that two cells of the same facies belong to the same connected component is higher in the training grid than in the simulation grid. This is consistent with the difference of grid dimensions. When the grid dimension along the channel di-

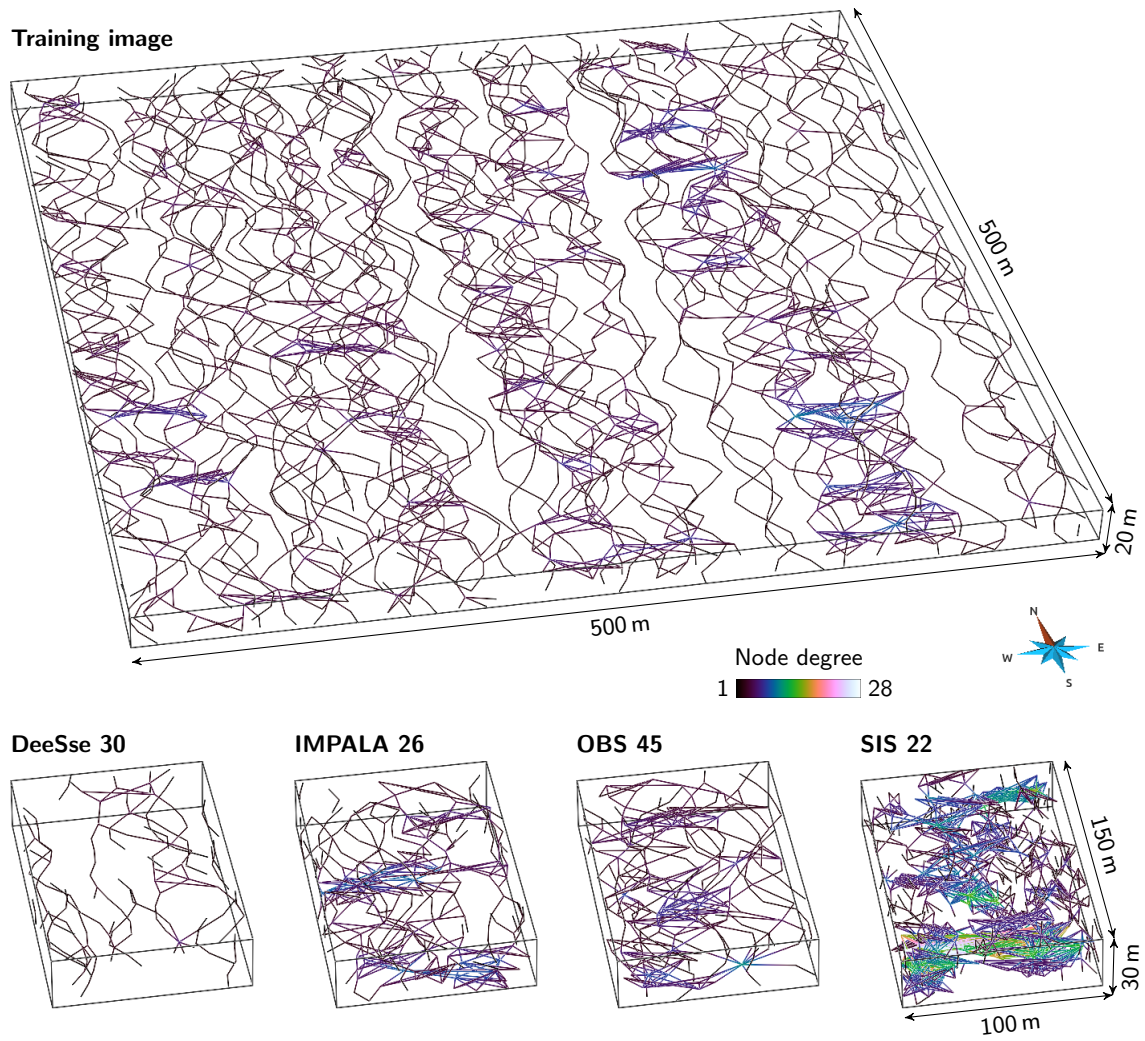


Figure 6 View of all the skeletons of the connected components for the TI and for a realization of each category.

rection increases, the probability that two channels cross each other to form a single connected component increases too, especially here with sinuous channels. In such case, the grid size impacts the characteristics of the connected components and the associated indicator values.

Comparing samples appear to be essential with grids of different dimensions. And using samples reveals other aspects of the images. For instances, the different samples coming from the TI are highly dissimilar. This illustrates the non-stationarity of the TI concerning the connectivity: some areas contain only one connected component as the channels are all connected, whereas other areas contain more connected bodies.

3.5 Comparing the connectivity of the training image and of the realizations

The purpose is now to compare the training image to all the realizations. These realizations come from different methods, but all borrow their input from the training image and have to reproduce the sedimentary bodies of the training image. All the following analysis relies on the image samples and not on the original images to avoid any bias due to the difference of size between the training image and the realizations.

3.5.1 Analysis of the dissimilarities

The dissimilarities give a first insight on the relationships between the different realizations (figure 10). The training image samples fall within the OBS samples, highlighting the similarity of these images. The samples from the multiple-point methods, DeeSse and IMPALA, are close to the OBS samples, but they do not mix up much. All these images are so not completely similar. Furthermore, the DeeSse and IMPALA samples remain away from the TI samples. The SIS samples are clearly distinct from all the other samples, and are the most distant from the TI samples.

If the confidence of the two-dimensional MDS representation is not high, the heat map confirms those observations (figure 11). The first row shows the dissimilarities between the training image samples and the realization samples. The whitest samples, the OBS ones, are the closest to the TI. The reddest samples, the SIS ones, are the furthest from the TI samples. The DeeSse and IMPALA samples fall in between, and seem equally close to the TI samples. Globally, the differences between all the methods are significant, as highlighted on the MDS.

As observed in the previous section, the training image samples are dissimilar one from the other. It shows the heterogeneity of static connectivity within the training image. Concerning the realizations, the OBS realizations are also dissimilar one from the other, whereas the SIS realizations are all really close.

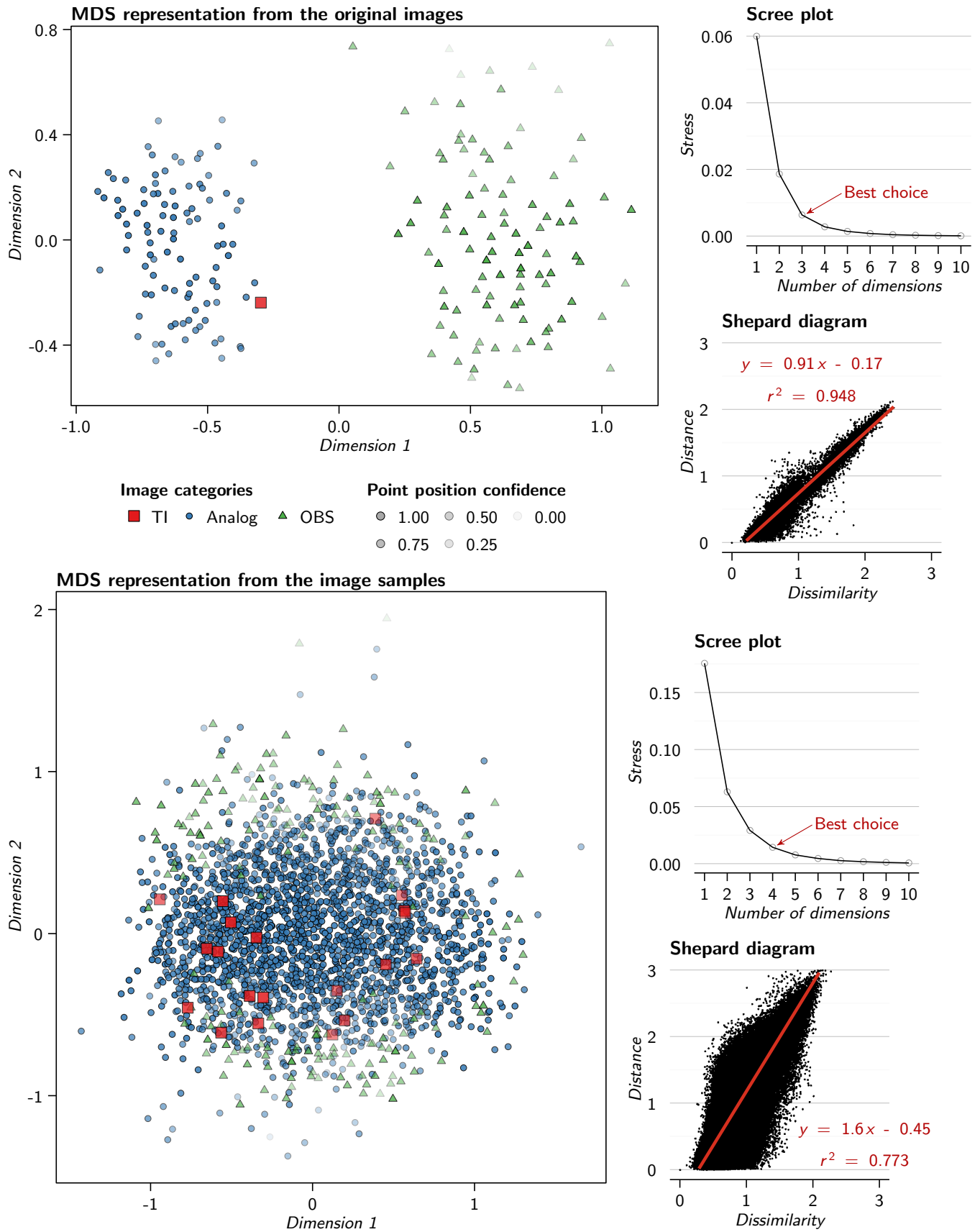


Figure 7 MDS representations of the dissimilarity matrices for the original images (with corrections of the indicators to cancel the effect of the grid dimensions) and samples (of same size). The scree plot for the original images only displays the stress values up to 10 dimensions on 200 possible. The scree plot for the samples only displays the stress values up to 10 dimensions on 2220 possible.

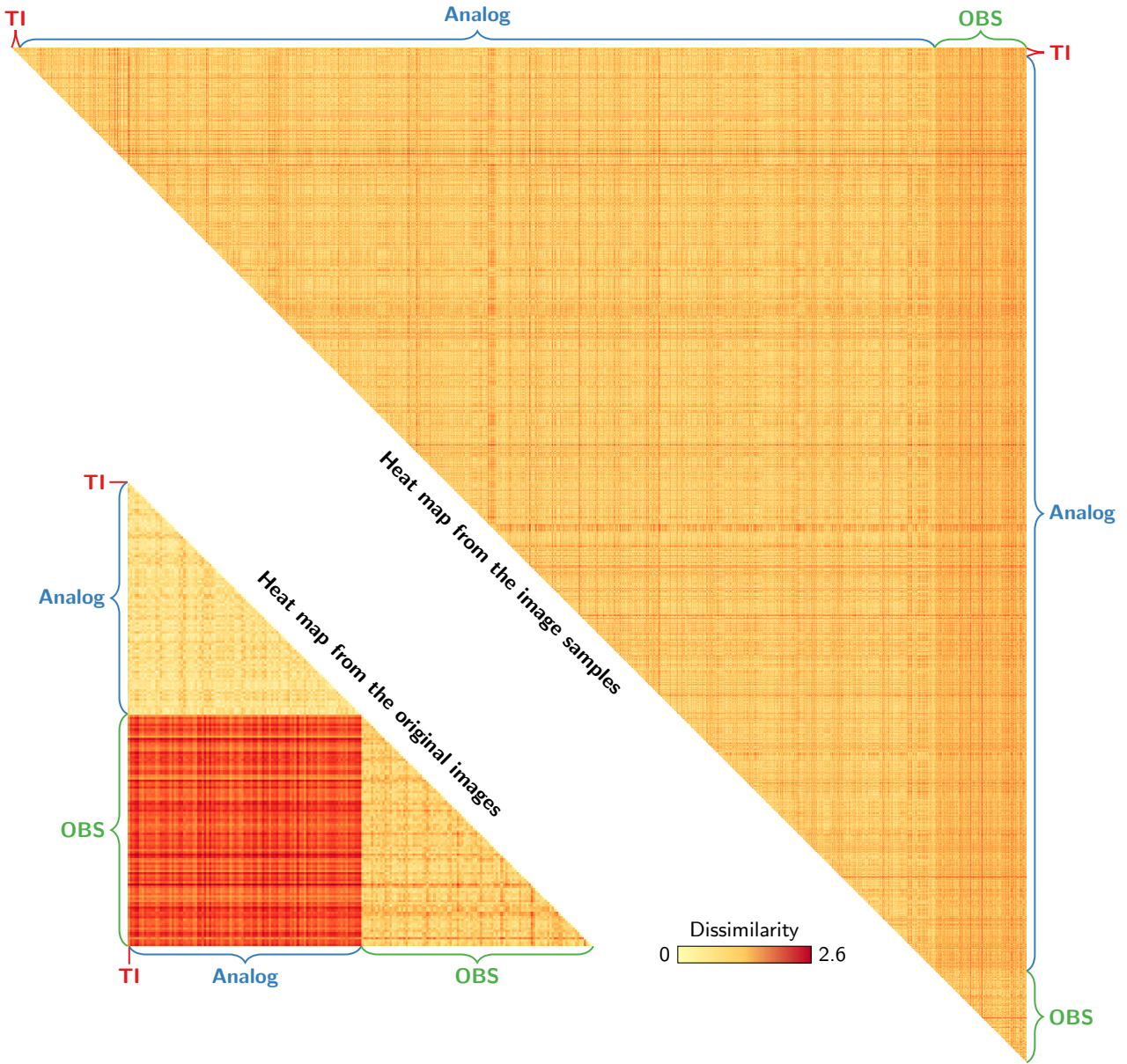


Figure 8 Heat map representations of the dissimilarity matrices for the original images (with corrections of the indicators to cancel the effect of the grid dimensions) and samples (of same size). Only one triangle of the symmetric matrices is represented.

Both DeeSse and IMPALA realizations are more spaced than the SIS ones, but not as much as the OBS realizations. All this tends toward a variable diversity concerning the static connectivity for the different methods in that case study. Going back to the indicators helps to further analyze such behavior.

3.5.2 Analysis of the indicators

The indicator values for the channels (figure 12 and 14) and levees (figure 13 and 15) differ depending on the category. The differences are more or less clear depending on the indicator, whose behavior differs between the two sedimentary body types.

The OBS samples being similar to the training image samples appear also on the indicator values. These values are close – and for many indicators the closest – to the TI values for the channels. That trend is less obvious with the levees, with less close values. But the levee density is the only indicator to be really away from the TI values. All this confirms the close relationship between the training image and the OBS

realizations concerning the static connectivity. It also confirms the visual observations. This is consistent with the use of the same method and parameters to simulate the training image and the OBS realizations.

Similarly, the significant dissimilarity between the SIS samples and the TI samples also appears on the indicator values. This is obvious on the traversing component proportion or the component density. The high component density means a higher number of connected components compared to the other samples. On the other side, the average number of component cells is quite low, meaning that most of these numerous components are small. The low traversing component proportion signifies that most of these components are not continuous enough to represent channels nor levees. Concerning channels, the significant difference between the SIS and TI samples for the shape indicators – number of component cells, box ratio, faces/cells ratio and sphericity – implies that the SIS components do not look like channels. This different shape also appears on the node degrees, with far higher node degree values than for the other categories, implying a less linear

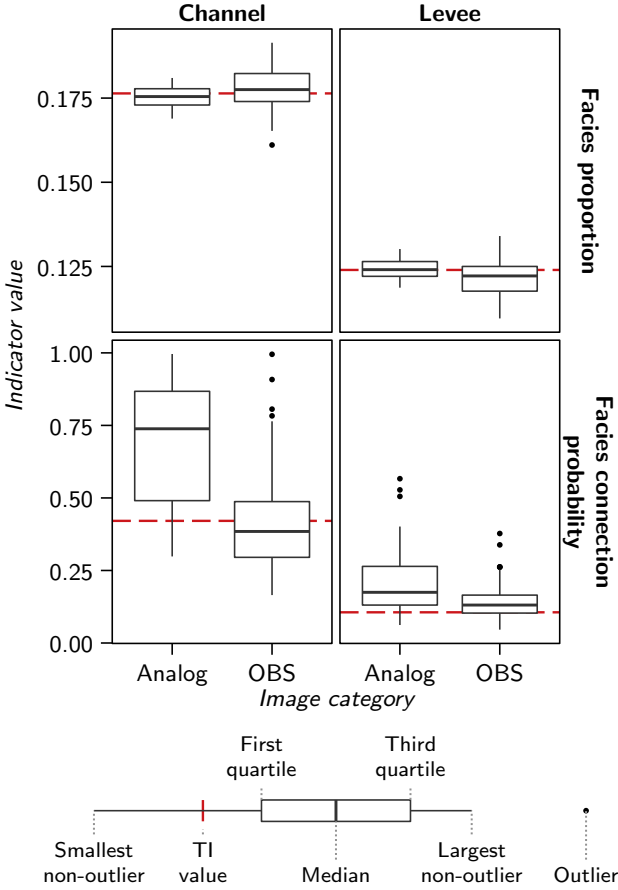


Figure 9 Box-plots comparing the facies proportions and facies connection probability for the TI, some TI analogs and the OBS realizations.

shape. Despite numerous and small components, the channel connection probability remains high. This means that these samples must contain one large component. This component must be traversing, as the traversing component proportion is not equal to zero. All those observations are consistent with the visual inspection of the realization, and confirm a significant difference of static connectivity between the training image and the SIS realizations. Many indicators also display a narrow range of values. This confirms the low variability between the SIS samples concerning the connectivity, as seen with the dissimilarities.

DeeSse and IMPALA samples have similarities with the SIS samples, especially more, and smaller, connected components than in the training image, as visible on the component density and the number of component cells. Similarly, the shape indicators show a significant difference between the TI samples and both the DeeSse and IMPALA samples. The higher sphericity implies in particular less linear shapes for the channels. Despite being equally dissimilar to the TI samples, the other indicators show significant differences between the DeeSse and IMPALA samples and the TI samples. The DeeSse samples have far lower channel and levee proportions. This impacts the facies adjacency, with channels and levees being more adjacent to the mudstone. But the most relevant difference between the DeeSse and IMPALA samples comes from the channel connection probability: the connection probability of the DeeSse samples is lower than that of the TI samples, whereas the connection probability of the IMPALA samples is higher than that of the TI samples. The IMPALA samples have a behavior sim-

ilar to that of the SIS, with a few large component among smaller ones. But these component connectivity is not completely similar to that of the SIS. This is especially visible on the node degree proportions, with the IMPALA samples having an intermediary behavior between the TI and the SIS samples. On the other side, the DeeSse sample connectivity seems lower. The higher degree two proportion of the DeeSse samples implies few intersections between channels. The higher degree one proportion also implies more discontinuous components. Again, all of this is consistent with the visual observations: DeeSse channels are clearly identifiable but discontinuous, whereas IMPALA channels are less visible, with many intersections.

In this case, the indicators confirm what comes from the dissimilarities: the OBS realizations are the most similar to the training image from a static connectivity perspective. This is consistent with the visual observations, and with the use of the same method to simulate the TI and the OBS realizations. The next section endeavors to compare those results from what can be obtained with multiple-point histograms.

3.6 Comparison with multiple-point histograms

Multiple-point histograms or pattern histograms have made their way as indicators of a realization quality with MPS methods [e.g., Boisvert et al., 2010, Tan et al., 2014]. We propose here to compare the results obtained with those histograms to the previous results. The histograms are based on a $3 \times 3 \times 3$ pattern and are computed on three levels of multi-grids [Tran, 1994], giving three histograms per image. The dissimilarity δ between two images r and s is adapted from the work of Tan et al. [2014]:

$$\delta(H^r, H^s) = \sum_{l=1}^3 \frac{1}{2^l} D_{JS}(H_l^r, H_l^s) \quad (11)$$

with H^r and H^s the sets of three histograms for each image, l the multi-grid level and D_{JS} the Jensen–Shannon divergence, which is the squared Jensen–Shannon distance. A multi-grid level l of 1 corresponds to the finest level and here 3 is the coarsest level. The coarser levels characterize the large-scale behavior of the sedimentary bodies. But they induce a loss of information. This justifies the decreasing weights when the multi-grid level increases. Similarly to the work using multiple-point histograms, the comparison is directly made on the original images, not on samples.

The observations about the category relationships made with the previous indicators (figure 10) remain valid on the MDS representation from the multiple-point histogram (figure 16). The training image falls within the OBS realizations. The DeeSse and IMPALA realizations are close from the OBS ones, but with a clear separation. They all remain separated from the TI. Again, the SIS realizations are far away from all the other images, including the TI. The main difference with the previous indicators comes from the variability within a category. This is especially noticeable with the SIS realizations, which seem to have a significant pattern variability.

The two-dimensional MDS representation is here again a poor representation of the dissimilarities, with a high stress. Only the dissimilarities with the training image are kept to directly study them and compare the ranking between different indicators (figure 17). Looking at all the connected component indicators – i.e., all the indicators described in table 1 – points

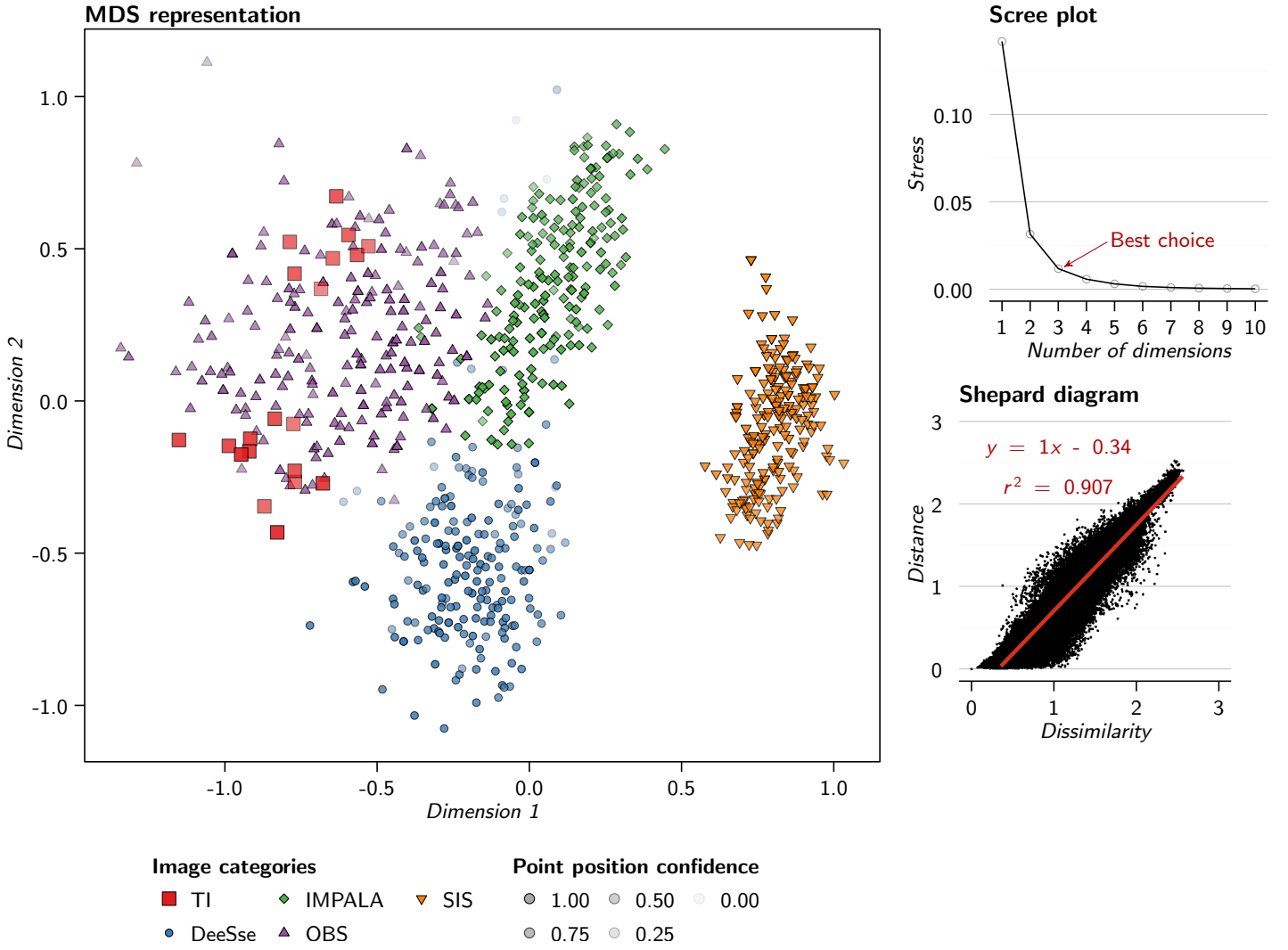


Figure 10 MDS representation of the dissimilarities between the samples of the case study generated using SMACOF and validation graphs. The scree plot only displays the stress values up to 10 dimensions on 820 possible.

out the conclusions coming from figure 10: the OBS realizations are the closest to the TI, the SIS ones the furthest, and the DeeSse and IMPALA realizations stand in between. Similar rankings come from the shape indicators – i.e., number of component cells, box ratio, faces/cells ratio and sphericity – and the skeleton indicators – i.e., node degree proportions and inverse branch tortuosity.

The multiple-point histograms have also a similar ranking, with a clearer separation between the SIS realizations and the other realizations (figure 17, All the multi-grid levels). However, the dissimilarities between the training image and the DeeSse realizations vary significantly between the multi-grid levels. The largest multi-grid level even places the DeeSse realizations closer to the TI than the OBS realizations. This level characterizes the large-scale behavior of the sedimentary bodies. Such ranking is then particularly surprising due to the presence of discontinuous bodies within the DeeSse realizations, but neither within the OBS ones nor within the TI. These continuity differences are confirmed by the skeletons, especially the higher proportion of node of degree one inside the grid for the DeeSse than for the OBS realizations.

4 DISCUSSION

The previous section highlights the ability of the method to distinguish realizations by focusing on the static connectivity through the connected components. This section discusses some aspects of the analysis process.

4.1 About the indicators

All the indicators proposed here rely more or less directly on the connected components. Some of them are classical, such as the facies proportion, but as highlighted on figure 9 the facies proportion is not enough to characterize the static connectivity. New indicators are introduced here compared to previous studies on connected components [Deutsch, 1998, De Iaco and Maggio, 2011]. Some indicators lead to better characterize the component organization, such as the traversing component proportion or the component density. Other indicators aim to better characterize the component shape, such as the sphericity. Using skeletons is also a new feature to compare realizations. The node degree proportion appears to give many details about the connectivity. The branch tortuosity has been less useful for the studied case, with a poorer discrimination of the realizations. This is due to the parameterization of the skeletonization, which favors the topology at the cost of the geometry of the skeletons.

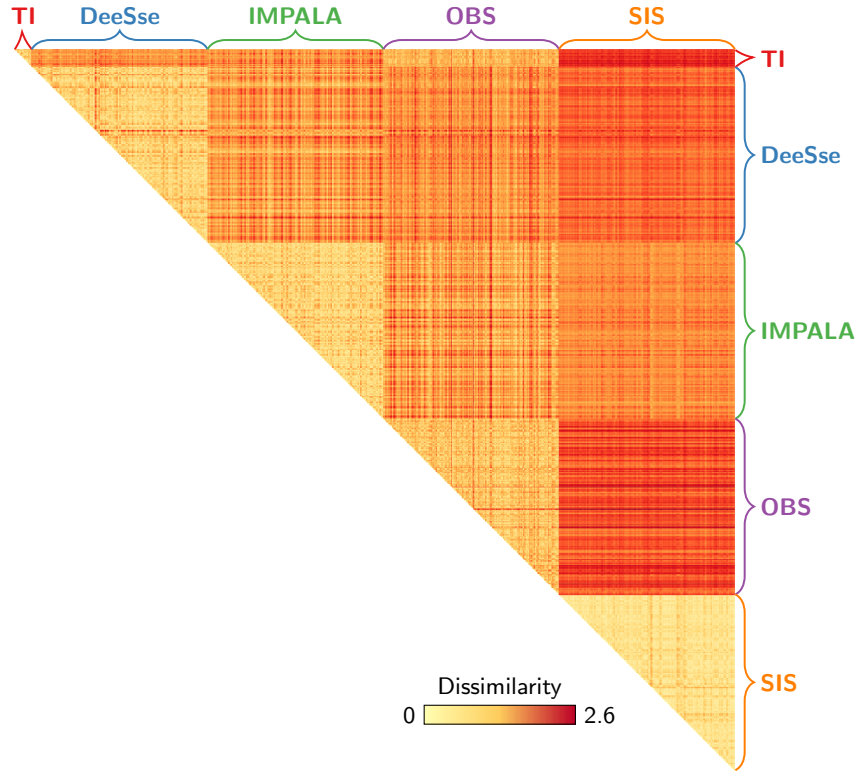


Figure 11 Heat map representation of the dissimilarity matrix computed based on the samples of the case study.

The use of multiple-point histograms as indicators in a method similar to Tan et al. [2014] shows a ranking close to that with the connected component indicators. However, they do not characterize the realizations in the same way. The multiple-point histograms of the finest multi-grid or multi-resolution level characterize in details the shape of the sedimentary bodies. The shape indicators are global measures over a whole connected component. As connected components can have variable shapes due to the sedimentary bodies intersections, being able to characterize more finely the component shape is an interesting asset. From this point of view, the multiple-point histograms could bring further information on the connected component shape.

However, the multiple-point histograms do not measure the static connectivity: they compare the patterns between the images, but not really the relationships between the patterns. The study of the coarsest multi-grid or multi-resolution levels attempts to look at the large scale behavior of the sedimentary bodies. But many details are lost in the process, what justifies the lower weights for these levels in the dissimilarity from multiple-point histograms [Tan et al., 2014]. And it still not characterizes the static connectivity. From this point of view, the skeletons describe more precisely the large-scale behavior of the components and their connectivity.

4.2 About indicator comparison

As stated in the previous section, a single indicator is not enough to fully characterize the static connectivity. Comparing several indicators lead to more relevant information about the realizations and how much they differ from the viewpoint of connectivity. Comparing realizations on grid of different dimensions leads to issues non-addressed by previous studies [Deutsch, 1998, De Iaco and Maggio, 2011]. A correction on the two most affected indicators is not sufficient to compensate

for different grid dimensions. Sampling the images appears to be more efficient, and also helps to analyze the connectivity heterogeneity within the images. The question of the sampling representativeness remains to be explored.

Using a metric is very useful, because it gathers all the indicator values into one dissimilarity value and facilitates the comparison of the realizations and the analysis. Tan et al. [2014] already used such process with multiple-point histograms. We have applied a similar principle to connected components, gathering many indicators into values easier to analyze. The introduction of a heterogeneous metric gives the opportunity to gather indicators of different types and further improves the method ability to characterize the realization static connectivity. At the end, the dissimilarities distinguish the realizations from different methods and parameter values, but also characterize the static connectivity variability between the realizations of a given method and parameter values.

Adding weights to the indicators in the metric computation means more flexibility for the user. Indeed, not all the indicators are significant to all the applications. For instance with a flow simulation purpose, the unit component proportion is not necessarily significant due to a fewer impact of the unit-volume component on the flow than channels. But such weights remain optional. In the case study, we did not discriminate the indicators with weights, because we wanted to study the information provided by all the indicators on the realizations. Studying the indicator values after the dissimilarities remains essential to better understand the static connectivity of the realizations.

4.3 About the skeletonization method

Skeletons enable to better characterize both the geometry and the topology of connected components. However, the skeletonization method influences both the geometry and the topol-

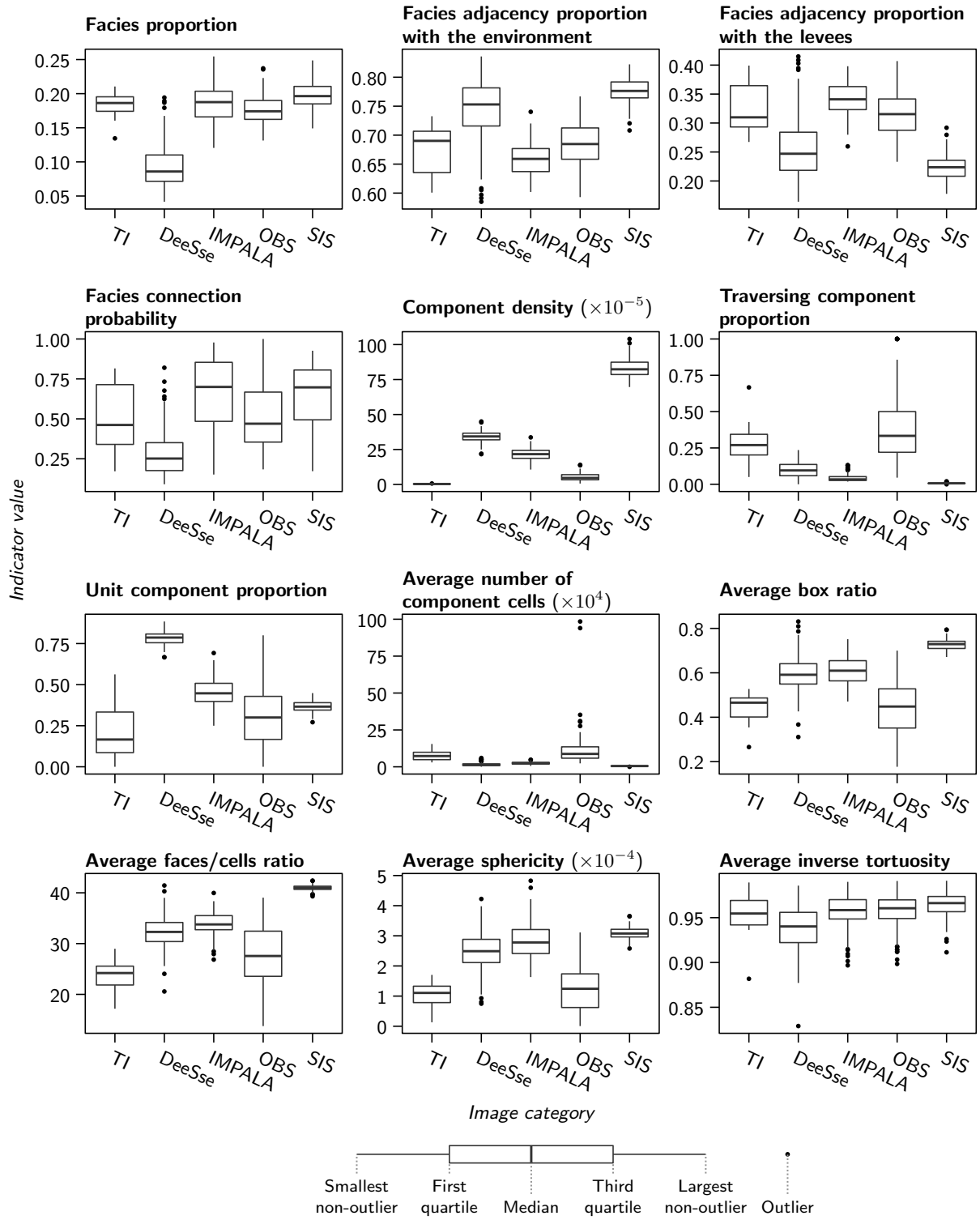


Figure 12 Box-plots comparing the range of indicators computed on the channels for the different categories, except the node degree proportions.

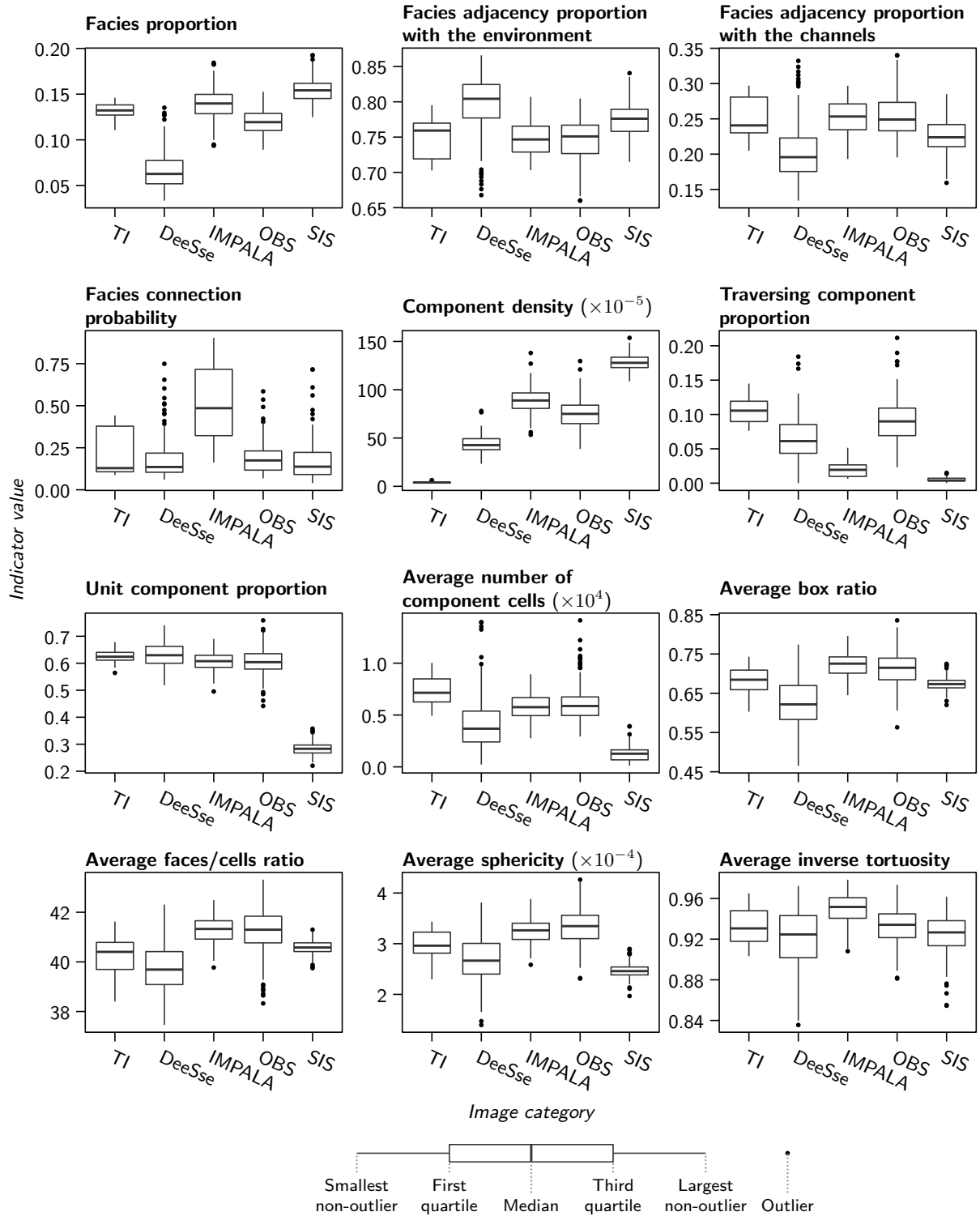


Figure 13 Box-plots comparing the range of indicators computed on the levees for the different categories, except the node degree proportions.

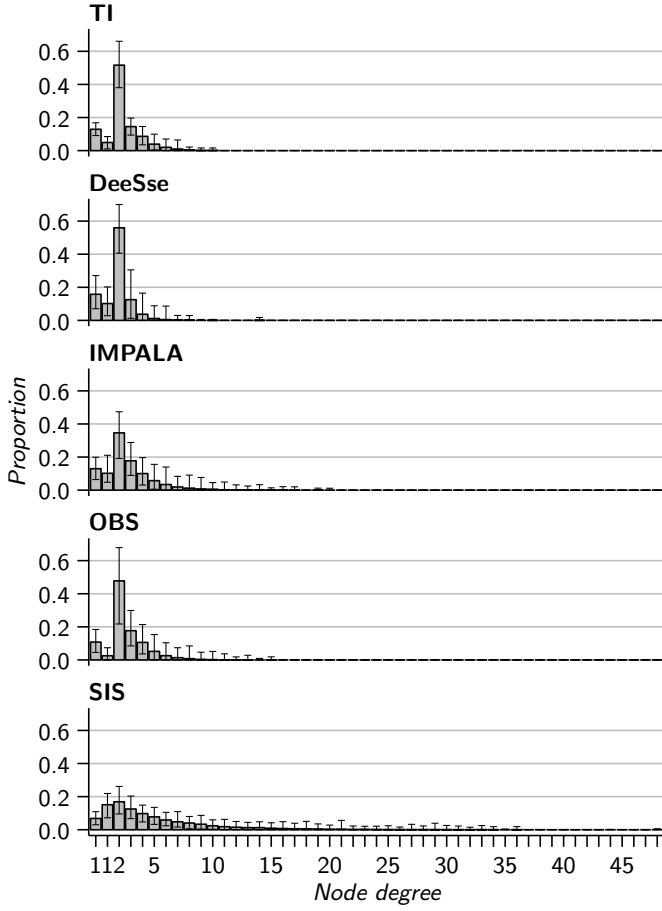


Figure 14 Mean node degree proportions of the channel skeletons for each category. The error bars display the minimum and maximum proportions. The first node degree 1 corresponds to the nodes of degree one along a grid border. The second node degree 1 corresponds to the nodes of degree one inside the grid.

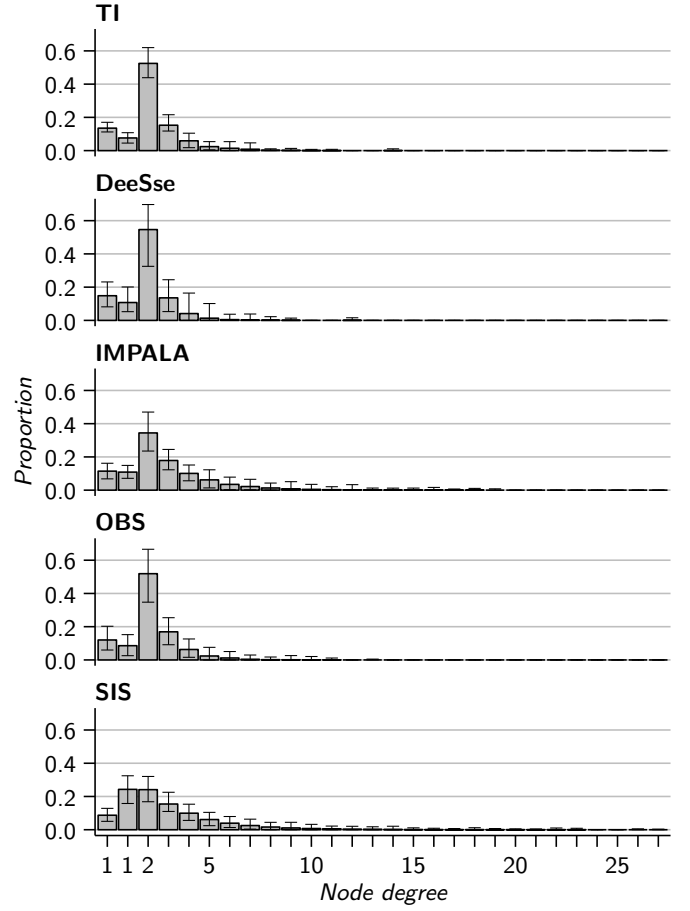


Figure 15 Mean node degree proportions of the levee skeletons for each category. The error bars display the minimum and maximum proportions. The first node degree 1 corresponds to the nodes of degree one along a grid border. The second node degree 1 corresponds to the nodes of degree one inside the grid.

ogy of the resulting skeletons. Among all the skeletonization methods, [Cornea et al. \[2007\]](#) distinguish the thinning-based method as the method with the best control on the skeleton connectivity. This section aims at comparing the result of a thinning-based method with the method introduced in section 1.3 based on slicing the grid and computing the connected components, denoted as the slicing-based method. The slicing-based method used hereafter is the algorithm defined by [Lee et al. \[1994\]](#) and implemented in the geomodeling software Gocad by [Barthélemy and Collon-Drouaillet \[2013\]](#).

The thinning-based method appears to perform better in two dimensions than the slicing-based method. But in three dimensions it tends to generate many small-scale loops (figure 18) which perturb both the topology and the geometry of the skeletons. The primary goal of the skeletons is to better characterize the large-scale topology – and possibly the geometry – of the connected components. The skeletons from the thinning-based method seem too perturbed to help in that characterization. The slicing-based method on the other side does not necessarily capture those small-scale elements due to the slice size. A large slice size may not capture the small components or all the component irregularities, but this is compensated in some way by the other indicators, in particular the shape indicators. Moreover, the thinning-based method tends to generate skeletons with many nodes, which are heavy to manipulate. The slicing-based method does not have the same issue when using quite high slice thicknesses. This aspect can

be essential when dealing with several hundreds of images.

All this leads to favor the slicing-based method in this work. Some aspects still need to be explored, such as the impact of the slice size. But many more skeletonization methods exist, even if skeletonizing three-dimensional shapes is an open debate. Further work could be done to study other methods and the topology and geometry of resulting skeletons.

4.4 About MDS methods and accuracy

We rely on the Scaling by MAjorizing a COmplicated Function as multidimensional scaling method to represent the dissimilarities. The SMACOF significantly facilitates the dissimilarity analysis. However, the dimensionality reduction makes the MDS representations imprecise, and the distances between the points tend to differ from the dissimilarities.

Thus, the MDS is not a simple visualization tool and can impact the analysis. This can be illustrated by comparing the MDS representation from the classical scaling (figure 19) and that from the SMACOF (figure 20) to analyze the dissimilarities between the original realizations and the original training image (and not the samples). Normally, the TI should stand from the realizations (see figure 7). But the classical scaling puts the TI close from the OBS and IMPALA realizations. Only the point position confidence shows that the TI position is wrong on the representation. The SMACOF representation separates more clearly the TI from the other images.

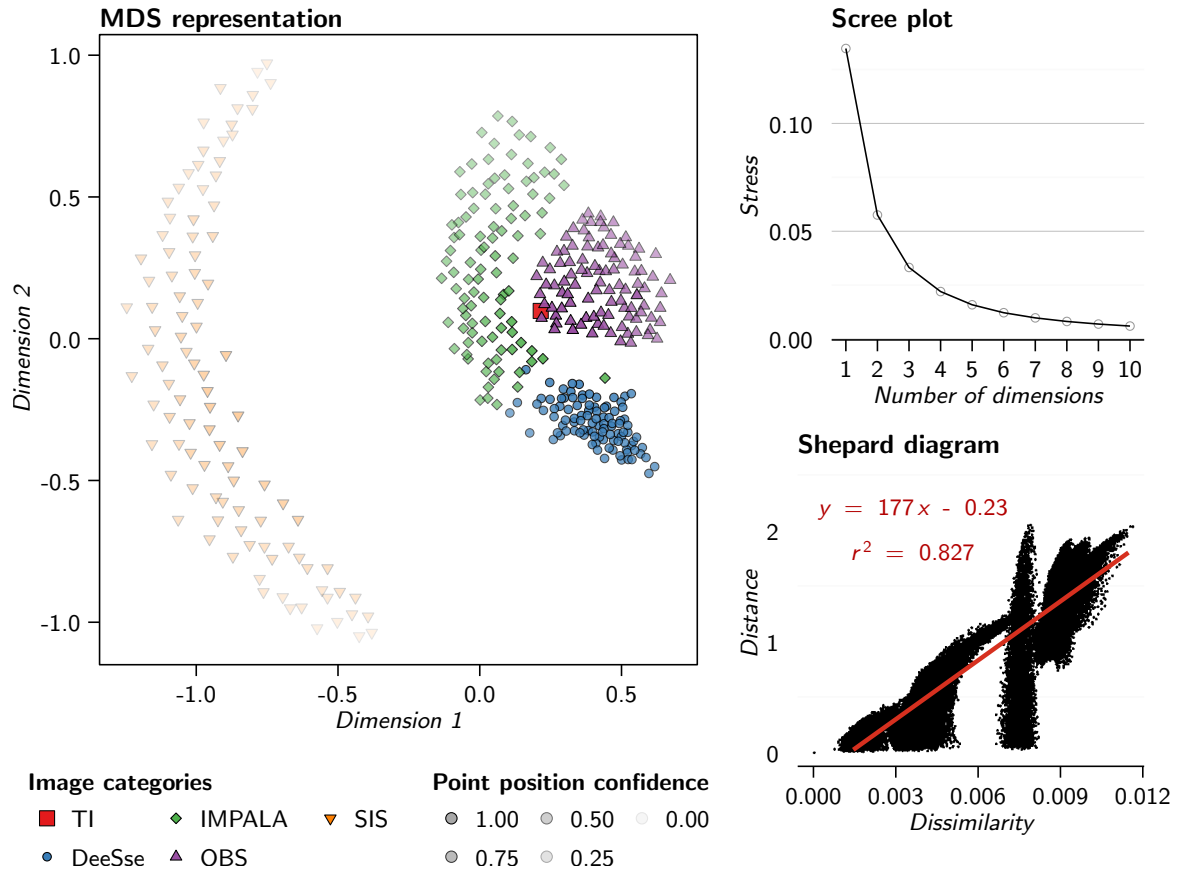


Figure 16 MDS representation of the dissimilarities between the images of the case study generated using SMACOF and validation graphs. The dissimilarities are based on the multiple-point (MP) histograms of the images. The scree plot only displays the stress values up to 10 dimensions on 400 possible.

Moreover, if the global relationships between the realization categories are similar between the two representations, the relative position of the images can be significantly different. This is clear with the TI, but also with other images (table 3). This appears more largely on the Shepard diagram, with a better coefficient of determination r^2 for the SMACOF than for the classical scaling. The classical scaling tends here to decrease the dissimilarities. As a result, the realization ranking can differ between the dissimilarities and the MDS distance (table 3). Thus, analyzing the sole MDS representation can lead to erroneous interpretations.

The choice of the MDS method is significant, so as the choice of the number of dimensions. We have privileged two-dimensional MDS representations for the sake of visibility, but three-dimensional representations would be worth testing. In any case, the MDS representation should always be cautiously studied and its misrepresentation of the dissimilarities should be kept in mind. From this point of view, the heat map facilitates the analysis of the real dissimilarity values. Analyzing a single row or column of the dissimilarity matrix – so comparing an image with all the others – is as easy to analyze as a MDS representation, but only on a subset of the images.

As the MDS facilitates the dissimilarity analysis, the dissimilarity simply makes the indicator analysis easier. After looking the MDS representation, it is essential to go back to the dissimilarity values to validate the observations. Similarly, studying the indicator values validates the observations and helps to further understand the difference of connectivity between the images.

4.5 Impact of the connected components on the flow

Facies heterogeneity shapes the fluid flow in subsurface. Thus, the petrophysical property simulation, directly correlated to the facies modeling, constitutes a preliminary step to further simulate flows. Being able to assess the static connectivity at the beginning of the workflow could constitute a real advantage in term of resources and time. It is also a way to ensure a better geological consistency, which in itself allows a better integration of field observations and measurements (seismic information, well data, etc.).

It would be interesting to apply the method on more detailed facies models than those of the case study. For instance, both channel and levee deposits are often heterogeneous, regardless of the sedimentary environment, with porous deposits more or less nested between flow barriers [e.g., Hubbard et al., 2011, Hansen et al., 2015]. Such flow barriers take the shape of mudstone drapes along the channel margins, of margin failure deposits, of channel abandonment deposits, etc. They can have a significant impact of the fluid flow [e.g., Labourdette et al., 2006, Pranter et al., 2007, Alpak et al., 2013, Issautier et al., 2013] and on the aquifer compartmentalization, with sometimes important consequences when they are ignored [see for instance Gainski et al., 2010, in a oil exploitation context]. Our method could clearly help to distinguish between several images from their differences in static connectivity – such images including realizations from different methods and/or parameter values or referring models. From this perspective, the case study shows the method ability to identify the simulation methods that produce subsurface models consistent with the static connectivity of a referring model.

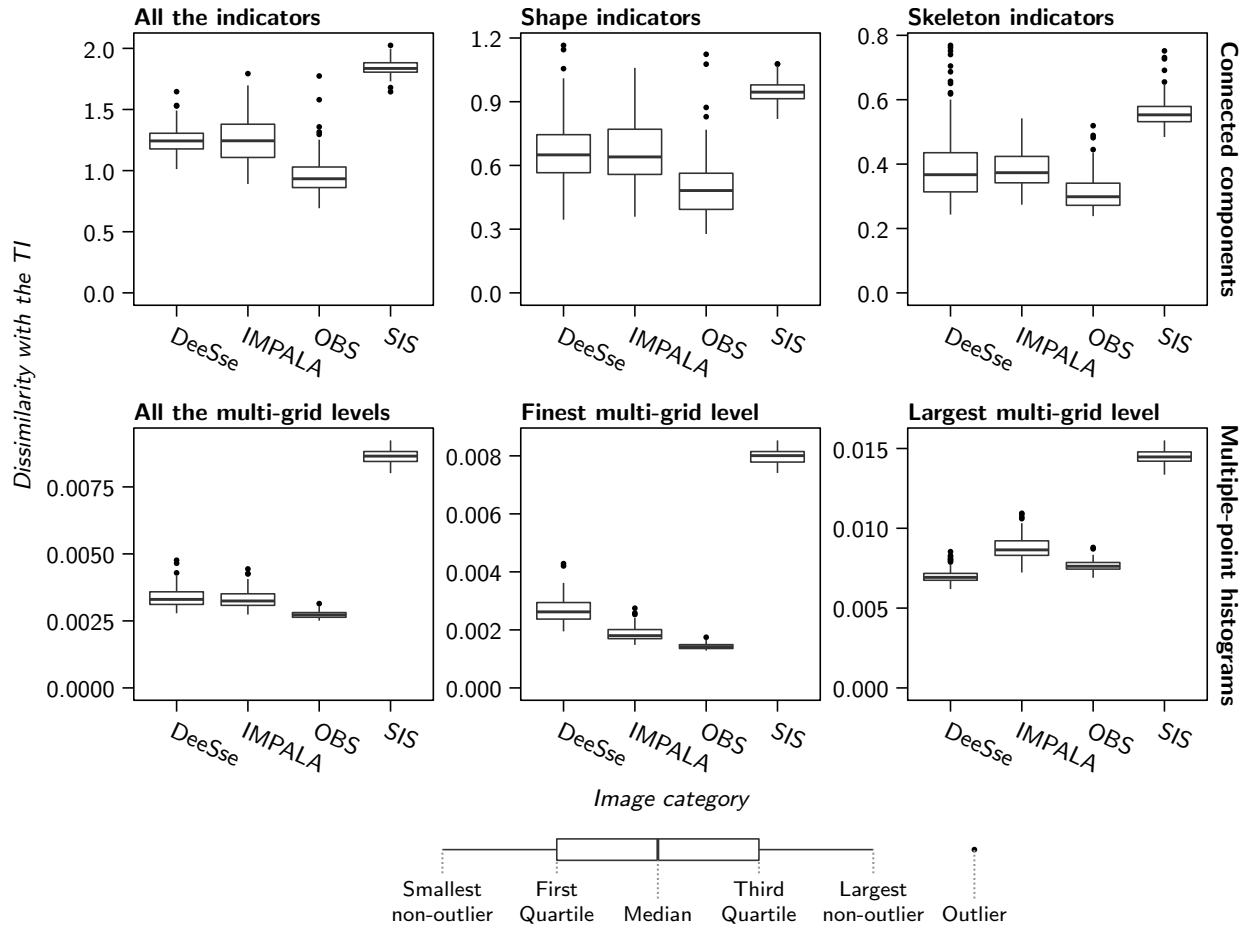


Figure 17 Box-plots comparing the realizations of each method with the TI. The dissimilarities depend on different indicators in each box-plot. Only the multiple-point histograms compare the original image directly, with one dissimilarity value with the TI per image. The other categories are based on image samples, and have several dissimilarity values with the TI. These values are averaged to obtain a single dissimilarity value with the TI.

Table 3 Comparison of dissimilarities and distances between the TI and DeeSse realization 12 and 76 for to MDS methods.

Compared images	Dissimilarity	Classical scaling distance	SMACOF distance
TI - DeeSse 12	1.579	0.596	1.414
TI - DeeSse 76	1.358	0.854	1.265
DeeSse 12 - DeeSse 76	0.905	0.259	0.203

Such approach is particularly adapted for fluvial and turbiditic channelized environments where channels tend to form high connectivity corridors, leading to channelized flow path. However, the static connectivity of a sedimentary body is not always representative of the flow behavior. For example, flow channeling can also emerge from non-channelized but highly heterogeneous bodies [e.g., [Park et al., 2008](#), [Fiori and Jankovic, 2012](#)]. It highlights the dependence of the hydrodynamic connectivity on many parameters: the permeability contrasts between the different media, the internal heterogeneity of each media, etc. Flow simulations then require to assign the petrophysical properties to each facies – usually with geostatistical methods [e.g., [Deutsch and Journel, 1992](#)]. Our metrics obviously do not anticipate the results of such procedures and, thus, just measure the consistency of the facies simulations in term of static connectivity.

Depending on the studied environment, the reproduction of the static connectivity could be secondary and one could directly work on hydraulic connectivity through the correspond-

ing properties. If reproducing the static connectivity does not guarantee to reproduce the exact hydraulic connectivity, it remains a step toward a better integration of geological information and knowledge in the physical description of the media. Our method provides a simple and objective basis for the comparison of large sets of realizations from this static connectivity point of view.

CONCLUSIONS AND PERSPECTIVES

This work develops an analysis process to compare several categorical images in terms of static connectivity. The process relies on the computation of dissimilarities from a given set of indicators. The indicators rely on connected components. They cover various aspect of the realizations, from their facies proportions to the size of their components. Their analysis in the presented case highlights their ability to distinguish between realizations simulated with different methods.

The multidimensional scaling summarizes all this informa-

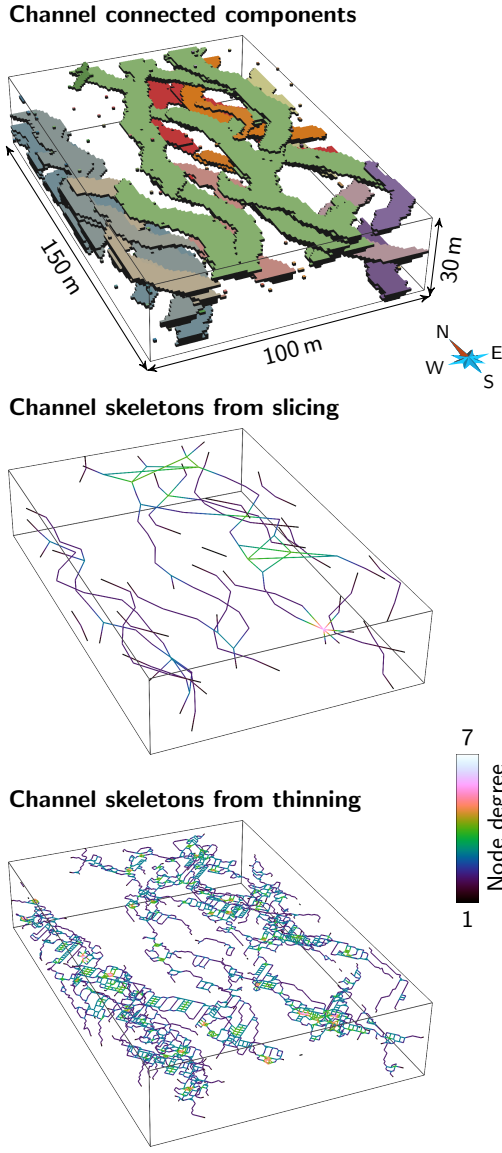


Figure 18 Skeletons from the channels of the realization DeeSse 30 generated with two methods: the slicing method used in the case study and the thinning method.

tion through low-dimensional representations of the dissimilarities. It is a powerful and practical visualization tool to get a first idea of the relationships between the images. But it should be kept in mind that MDS representations provide only a partial view of the dissimilarities. Moreover, the representation and its quality are method-dependent. The MDS representation should so always be interpreted cautiously. Analyzing the dissimilarity matrix shall always be favored rather than analyzing the MDS representations. The heat map helps to have an easier check on the MDS interpretation. Then an analysis should be done on particular rows or column of the matrix depending on the studied cases. This gives access to simple but powerful visualization tools that work on the real dissimilarities.

A detailed analysis shows the consistency of the indicator behaviors. Indicators of the connected component shape are interesting for their simplicity. But their sensitivity to the grid size is a real issue. A correction is proposed here, but it does not remove completely the error, and analyzing image samples is a better choice. Multiple-point histograms also seem a valid option to analyze more locally the component shape.

Skeleton-based indicators appear to be promising. The node degree proportions provide a simple way to compare connectivity structures. Further work should be done on the skeletons to better characterize both the connected component topology and geometry.

Thus, the present work shows that analyzing the behavior of various indicators needs to be continued. More case studies should be considered. Channels are known to have a heterogeneous filling, with shale acting as a flow barrier. It could be interesting to model and work on the component formed by the reservoir facies rather than by the sedimentary objects they belong to. Starting with real data would also be beneficial. If the proposed indicators focus on the static connectivity, an interesting perspective would be to explore their link with the dynamic connectivity.

ACKNOWLEDGEMENTS

This work was performed in the frame of the RING project at Université de Lorraine. We would like to thank the industrial and academic sponsors of the Gocad Research Consortium managed by ASGA for their support and Paradigm for providing the SKUA-GOCAD software and API. We also thank Sebastien Strebelle for its helpful remarks which influenced this research. We would like to thank the reviewers, including Aldo Fiori, for constructive comments which helped improve this paper.

APPENDIX A DETAILED DEFINITIONS OF THE INDICATORS

The following appendix describes more precisely all the indicators, starting with the indicators computed on the connected components and ending with those computed on the skeletons.

A.1 Global indicators

The global indicators do not characterize a particular connected component but an ensemble of connected components. They provide one value per facies.

A.1.1 Facies proportion

The facies proportion is of major importance in reservoir modeling considering its influence on porosity and permeability simulation. The proportion p of a facies f is defined as:

$$p_f = \frac{n_f}{n_t} \quad (12)$$

with n_f the number of cells of facies f and n_t the total number of cells.

A.1.2 Facies adjacency proportion

The adjacency proportion between the facies quantifies the spatial relationships between those facies. The proportion p^a of a facies f adjacent to a facies j is defined as:

$$p_{f,j}^a = \frac{n_{f,j}}{n_{f,t}} \quad (13)$$

with $n_{f,j}$ the number of cells of facies f adjacent to the facies j and $n_{f,t}$ the total number of cells of f adjacent to a facies different from f .

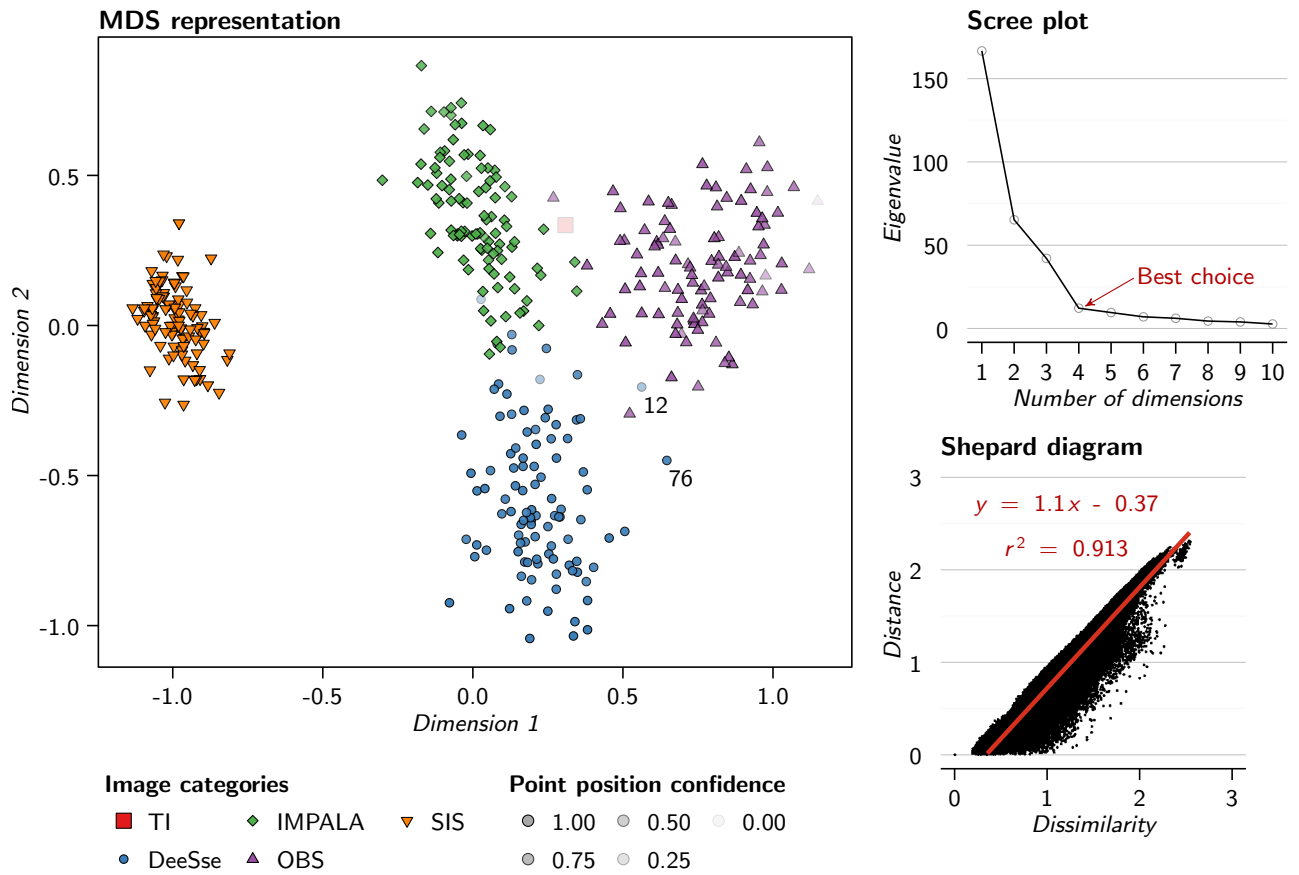


Figure 19 MDS representation of the dissimilarities between the images of the case study generated using classical scaling and validation graphs. The scree plot only displays the eigenvalues up to 10 dimensions on 400 possible.

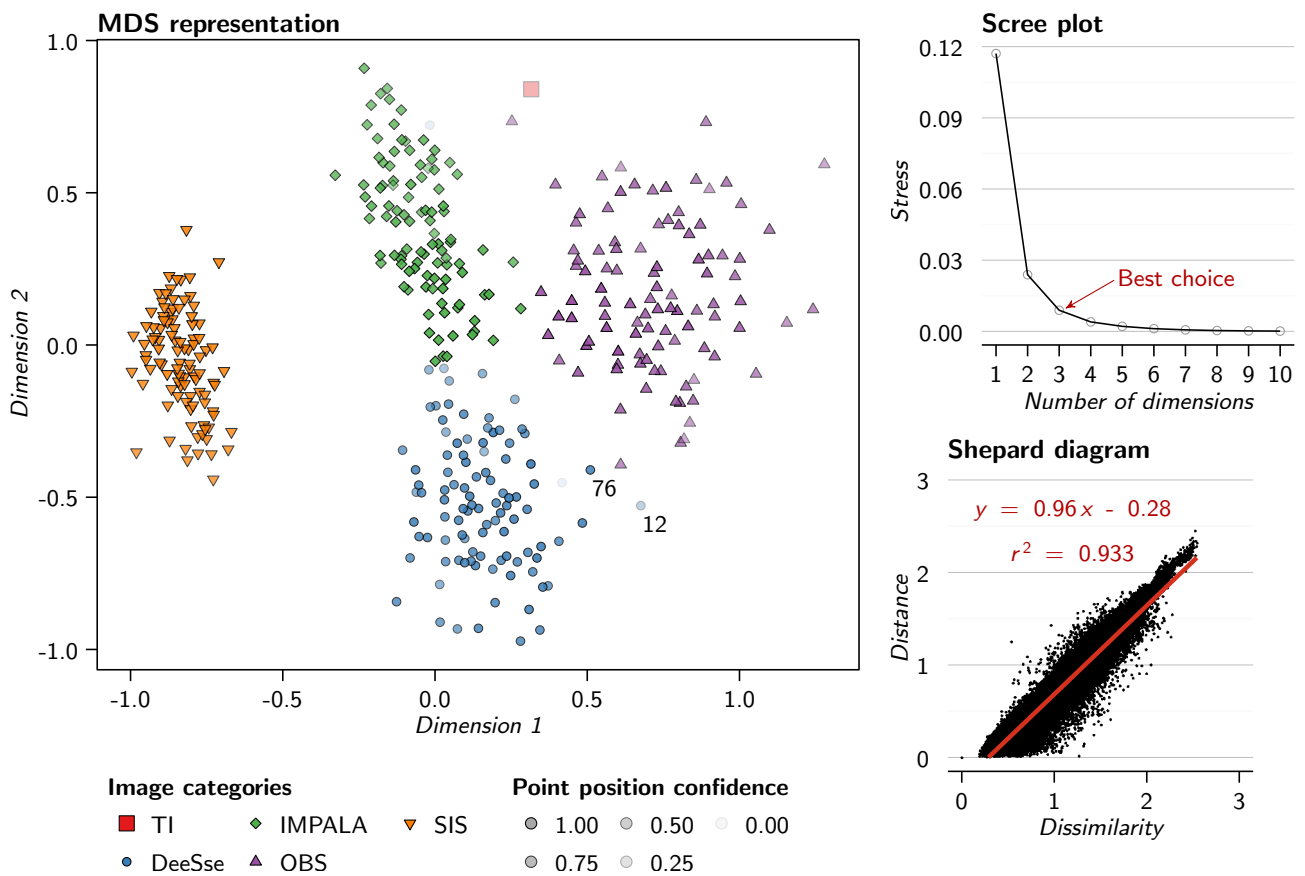


Figure 20 MDS representation of the dissimilarities between the images of the case study generated using SMACOF and validation graphs. The scree plot only displays the stress values up to 10 dimensions on 400 possible.

A.1.3 Facies connection probability

The connection probability Γ quantifies the global connectivity of a given facies f . It is calculated as the proportion of pairs of connected cells among all the pairs of cells of the considered facies [Renard and Allard, 2013]:

$$\Gamma_f = \frac{1}{n_f^2} \sum_{i=1}^{N_f} (n_f^i)^2 \quad (14)$$

with n_f the total number of cells of facies f , N_f the number of connected components of the facies f and n_f^i the number of cells of the connected component i associated to facies f .

A.1.4 Connected component density

The density ϵ of connected components for a facies f expresses the number of components per image cells:

$$\epsilon_f = \frac{N_f}{n_t} \quad (15)$$

with N_f the number of connected components of the facies f and n_t the total number of cells.

A.1.5 Unit connected component proportion

The unit connected component proportion quantifies the number of connected components constituted by only one cell. The creation of such bodies may correspond to some erroneous and unwanted small scale noise. In that case, it is better to avoid them for the shape indicator computation (see A.2). The proportion p^u of unit volume connected components of a facies f is:

$$p_f^u = \frac{N_f^u}{N_f} \quad (16)$$

with N_f^u the number of unit volume connected components of the facies f and N_f the number of connected components for the same facies.

A.1.6 Traversing connected component proportion

Connected components can have various behaviors against the grid: they can be in the middle of the grid without contact with a border (figure 2, connected component 3), along one border, along two adjacent borders (figure 2, connected component 2) or traversing the grid completely from one border to the opposite border (figure 2, connected component 1). These last components having a main impact on the flow, their reproduction is of prime importance. The proportion p^c of traversing connected components of facies f is defined as the proportion of connected components that connect one border to the opposite one:

$$p_f^c = \frac{N_f^c}{N_f - N_f^u} \quad (17)$$

with N_f^c the number of traversing components for the facies f , N_f the total number of components for the facies f and N_f^u the number of unit components for the facies f .

A.2 Shape indicators

Shape indicators all give one value per component. The arithmetic mean of those values provides a value of the indicator for a given facies. This makes the indicator comparison easier. The unit-volume connected components (see section A.1.5) give little information on the shape of the most important connected bodies considering flow circulation: the channels. But they can have a significant influence on the averaged shape indicators (section A.2). They are so ignored in the average computation.

A.2.1 Number of connected component cells

The number of cells n_f^g within a connected component g of a facies f represents the size of a connected component.

A.2.2 Box ratio

The box ratio is based on the axis-aligned bounding boxes of the components (figure A.1). It is related to the tortuosity of a component and to its orientation relative to the grid axis. The box ratio β of the connected component g of the facies f is expressed between 0 and 1 using the following formula:

$$\beta_f^g = \frac{n_f^g}{n_f^b} \quad (18)$$

with n_f^g the number of cells of the connected component g of the facies f and n_f^b the number of cells occupied by its axis-aligned bounding box.

A.2.3 Faces/cells ratio

The faces/cells ratio is equivalent to the surface/volume ratio, which compares the surface area of a component with its volume. Deutsch [1998] uses it as a measure of the tortuosity of the components, but it is also affected by their roughness. The faces/cells ratio ζ of the connected component g of the facies f is expressed as:

$$\zeta_f^g = \frac{m_f^g}{n_f^g} \quad (19)$$

with m_f^g the number of faces along the surface of the connected component g of the facies f and n_f^g its number of cells.

A.2.4 Sphericity

The sphericity ϕ compares the shape of a connected component g for a facies f with a sphere [Wadell, 1935]. It ranges between 0 and 1, where 1 corresponds to a sphere, and is expressed as:

$$\phi_f^g = 36\pi \frac{(n_f^g)^2}{(m_f^g)^3} \quad (20)$$

with n_f^g the number of cells of the connected component g of the facies f and m_f^g its number of faces along its surface.

A.2.5 Skeleton indicators

Skeletons are one-dimensional representations of the connected components, on which measures of geometry and topology are done.

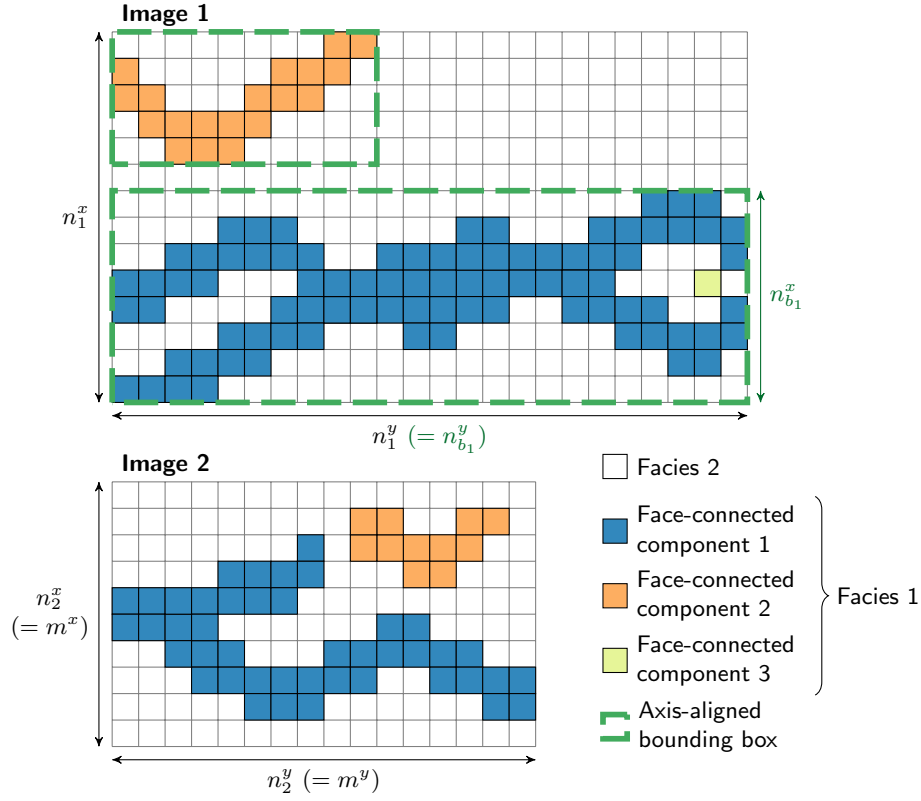


Figure A.1 Connected components of figure 2 represented with their bounding boxes (image 1). Image 2 is another image in a different grid. The two grids have different dimensions along the x axis, which can influence both the number of connected components and the number of grid cells. But they also have different dimensions along the y axis, along which the channels are oriented. Changes along the y axis influence the number of grid cells, but not the number of channels. In that case, the comparison of the density of the two images is biased. The same principle stands for the number of cells of each components. Differences between the number of cells for the component 1 of each image mainly come from the difference in grid dimension along y , not from real differences of connectivity. n_1^x and n_1^y represent the number of cells along the axis x and y for image 1, n_2^x and n_2^y the number of cells along the axis x and y for image 2 and $n_{b_1}^x$ and $n_{b_1}^y$ the number of cells along the axis x and y for the axis-aligned bounding box b of the connected component 1. m^x and m^y are the lower sizes between the two grids for the axis x and y (see appendix B for more details).

A.2.6 Node degree proportion

The node degree proportion is based on the number of segments connected to a given node (figure A.2). It is related to the topology of the skeleton and gives access to a deeper study of the connectivity of the structures of interest. The proportion p^n of nodes of degree n over all the skeletons for a given facies f is expressed as:

$$p_f^n = \frac{o_f^n}{o_f} \quad (21)$$

with o_f^n the number of nodes of degree n for all the skeletons of the connected components of the facies f and o_f the total number of nodes of the skeletons of the connected components for the facies f . Here we separate into two different classes the nodes of degree one along a border of an image and those in the middle of the image. The first ones relate to component terminations due to the limited size of the image. The other ones may relate to a bad body reproduction, for instance with channels that should be continuous and without dead-ends. The proportions for the different degrees give a histogram of node degrees.

A.2.7 Inverse branch tortuosity

A branch is a part of a skeleton defined as an ensemble of segments that link nodes of degree 2 and delimited by two nodes

of degree different from 2 (figure A.2). The inverse tortuosity t quantifies how twisted a branch b is, with values ranged between 0 and 1. This measure is related to the geometry of the skeleton. It is expressed for a branch b belonging to a skeleton s of a component of a given facies as:

$$t_s^b = \frac{d_s^b}{l_s^b} \quad (22)$$

with l_s^b the curvilinear length of the branch b of the skeleton s and d_s^b the Euclidean distance between the two extremity nodes of the same branch. The values for all the branches of all the skeletons related to an image are averaged to obtain one indicator value for that image.

APPENDIX B INDICATOR CORRECTION FOR DIFFERENT GRID DIMENSIONS

The following appendix describes the corrections applied to the connected component density and the number of component cells when the analysis implies different grid dimensions.

B.1 Correction for the connected component density

When some connected components of a facies are traversing (see A.1.6), reducing the grid dimension along the component

Image 1

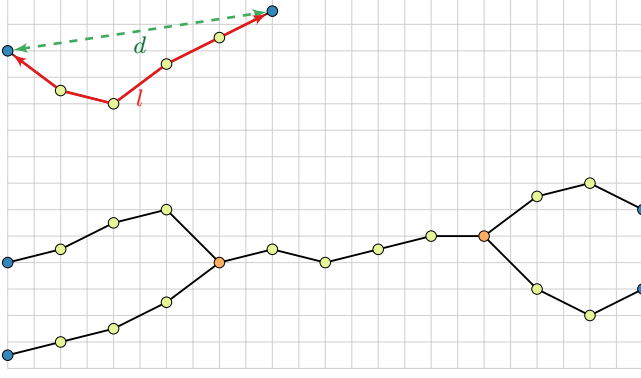


Image 2

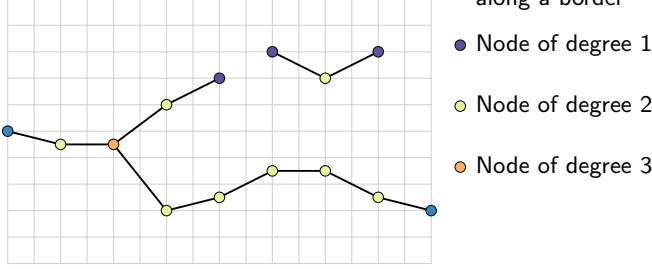


Figure A.2 Example of skeletons for the connected components of the figure A.1. If image 2 has no connection with a degree higher than 3, such as image 1, it displays some dead-ends highlighted by nodes with a degree one in the middle of the grid. This difference emphasize mis-reproductions of the channel connectivity. In image 1, l represents the curvilinear length of the branch and d the distance between its extremities. Those measures are used to compute the tortuosity, which can also differ between the images.

traversing extension can change the component size but not necessarily their number. For instance changing the grid size along the channel orientation does not change the number of channels. But it changes the grid volume and so does the density. Comparing two images of different sizes can affect the density whereas there is nothing wrong about the quality.

To overcome this aspect, the connected component density (formula 15) is replaced by:

$$\epsilon_f = \frac{N_f}{\vartheta_t} \quad (23)$$

with N_f the number of connected components of the facies f and ϑ_t is a corrected number of grid cells corresponding to a mix between the dimensions of each compared images:

$$\vartheta_t = \vartheta^x \times \vartheta^y \times \vartheta^z \quad (24)$$

with ϑ^x , ϑ^y and ϑ^z respectively the corrected number of cells along the axis x , y or z . Those corrected numbers of cells are chosen depending on the smallest axis size among all the compared images and the size of the connected components. For an ensemble S of images to be compared, we note m^i the minimal size along the axis i over the images in S :

$$m^i = \min_{s \in S} (n_s^i) \quad i \in \{x, y, z\} \quad (25)$$

with n_s^i the number of cells along the axis i for the image s . Determining the corrected numbers of axis cells for a given

image r of S depends on the size of its connected components:

$$\vartheta^i = \begin{cases} m^i & \text{if } \exists g \in r, n_g^i \geq m^i \\ n_r^i & \text{otherwise} \end{cases} \quad i \in \{x, y, z\} \quad (26)$$

with g a connected component of r , n_r^i the number of cells along the axis i for the image r and n_g^i the number of cells along the axis i for the component g .

For instance in (figure A.1), the component 1 of image 1 is larger than image 2 along the y axis. On the other hand, no component of that facies is larger than image 2 along the other axes. The corrected volumes for the image 1 $\vartheta_{t,1}$ and for the image 2 $\vartheta_{t,2}$ are so:

$$\begin{aligned} \vartheta_{t,1} &= n_1^x \times n_2^y \times n_1^z \\ \vartheta_{t,2} &= n_2^x \times n_2^y \times n_2^z \end{aligned} \quad (27)$$

with n_1^x , n_1^y and n_1^z the number of cells along the axis x , y and z for image 1 and n_2^x , n_2^y and n_2^z the number of cells along the axis x , y and z for image 2. Image 1 is so considered as truncated to limit the effect of the traversing bodies on the density.

B.2 Correction of the number of connected component cells

The objects are expected to have roughly the same range of sizes in all the images. But this expectation is no longer relevant in the case of traversing structures (see A.1.6), whose dimensions must vary following the size of the supporting grid. As for the connected geobody density, the possible difference of image sizes imposes to rescale the number of cells of a connected component.

The rescaling is based on the estimation of the component bounding box size if this component was transferred within the grid of the smallest size. That bounding box size can be defined following a number of cells ϑ_b :

$$\vartheta_b = \vartheta^x \times \vartheta^y \times \vartheta^z \quad (28)$$

with ϑ^x , ϑ^y and ϑ^z respectively the corrected numbers of cells of the bounding box along the axis x , y or z . Those corrected numbers of cells are following the same principles as to find the corrected number of cells for the density. Determining the corrected numbers of axis cells requires the same parameter m^i (formula 25):

$$\vartheta^i = \begin{cases} m^i & \text{if } n_b^i \geq m^i \\ n_b^i & \text{otherwise} \end{cases} \quad i \in \{x, y, z\} \quad (29)$$

with n_b^i the number of cells along the axis i for the bounding box of the component. Finally the corrected number of cells ϑ_f^g for the component g of the facies f is proportional to the ratio between the number of cells of its corrected bounding box ϑ_b and the number of cells of its actual bounding box n_b :

$$\vartheta_f^g = n_f^g \times \frac{\vartheta_b}{n_b} \quad (30)$$

For instance the component 1 on image 1 of figure A.1 has a corrected number of cells ϑ_1 of:

$$\vartheta_1 = n_1 \times \frac{n_{b_1}^x \times n_{b_1}^y \times n_{b_1}^z}{n_{b_1}^x \times n_{b_1}^y \times n_{b_1}^z} = n_1 \times \frac{n_2^y}{n_{b_1}^y} \quad (31)$$

with n_1 the number of cells of component 1, $n_{b_1}^x$, $n_{b_1}^y$ and $n_{b_1}^z$ the numbers of cells of its axis-aligned bounding box along the axis x , y or z and n_2^x , n_2^y and n_2^z the numbers of cells of image 2 along the axis x , y or z . This reflects the impossibility to make that component fit into image 2. Its size needs so to be rescaled to be compared with objects of image 2.

APPENDIX C SIMULATION PARAMETERS FOR THE REALIZATIONS

Table C.1 Parameters used to simulate the channelized environment with DeeSse.

Parameters	Values
Maximum number of neighbors	64
Acceptance threshold	0.05
Maximal scan fraction of the TI	0.33

Table C.2 Parameters used to simulate the channelized environment with IMPALA.

Parameters	Values
Number of multi-grids	4
Number of multi-grid levels in each direction	$4 \times 4 \times 1$
Search template type	Elliptic
Size of the search template (radii in m)	$7 \times 7 \times 4$
Maximal number of neighbors in the template	64

Table C.3 Parameters used to simulate the channelized environment with the object-based method of Petrel. The distributions used are all triangular.

Simulation parameters	Min	Mode	Max
Channels			
Proportion (in %)	21.21	21.21	21.21
Orientation (in °)	0	0	0
Amplitude (in m)	10	15	40
Wavelength (in m)	60	70	100
Width (in m)	7	10	13
Thickness (in m)	1.5	2	4
Levees			
Proportion (in %)	8.79	8.79	8.79
Width (in m)	4	7	11
Thickness (relative to channel thickness)	0.25	0.35	0.6

Table C.4 Variogram parameters used to simulate the channelized environment with SIS.

Variogram parameters	Channels	Levees	Mudstone environment
Azimuth (in °)	0	0	0
Dip (in °)	0	0	0
Sill	0.145	0.109	0.210
Nugget	0	0	0
Range 1 (in m)	23	26	70
Range 2 (in m)	12	14	34
Range 3 (in m)	3	1	2.5
Type	Spherical	Exponential	Exponential

REFERENCES

F. O. Alpak, M. D. Barton, and S. J. Naruk. The impact of fine-scale turbidite channel architecture on deep-water reservoir performance. *AAPG Bulletin*, 97(2):251–284, 2013. doi: 10.1306/04021211067. (Cited page 19)

C. Barthélemy and P. Collon-Drouaillet. Simulation of Anastomotic Karst Networks. In *Proc. 32nd Gocad Meeting*, page 13, Nancy, 2013. (Cited page 18)

J. B. Boisvert, M. J. Pyrcz, and C. V. Deutsch. Multiple Point Metrics to Assess Categorical Variable Models. *Natural Resources Research*, 19(3):165–175, 2010. doi: 10.1007/s11053-010-9120-2. (Cited pages 2 and 13)

J. W. Brandt and V. R. Algazi. Continuous skeleton computation by Voronoi diagram. *CVGIP: Image Understanding*, 55(3):329–338, 1992. doi: 10.1016/1049-9660(92)90030-7. (Cited page 3)

A. Comunian, P. Renard, and J. Straubhaar. 3D multiple-point statistics simulation using 2D training images. *Computers & Geosciences*, 40(0):49–65, 2012. doi: 10.1016/j.cageo.2011.07.009. (Cited page 2)

N. D. Cornea, D. Silver, and P. Min. Curve-skeleton properties, applications, and algorithms. *Visualization and Computer Graphics, IEEE Transactions on*, 13(3):530–548, 2007. doi: 10.1109/TVCG.2007.1002. (Cited page 18)

T. F. Cox and M. A. Cox. *Multidimensional scaling*. Chapman and Hall, London, 1994. ISBN 978-1-58488-094-3. (Cited page 6)

S. De Iaco. On the use of different metrics for assessing complex pattern reproductions. *Journal of Applied Statistics*, 40(4):808–822, 2013. doi: 10.1080/02664763.2012.754853. (Cited page 2)

S. De Iaco and S. Maggio. Validation Techniques for Geological Patterns Simulations Based on Variogram and Multiple-Point Statistics. *Mathematical Geosciences*, 43(4):483–500, 2011. doi: 10.1007/s11004-011-9326-9. (Cited pages 2, 14, and 15)

J. De Leeuw. Applications of convex analysis to multidimensional scaling. In *Recent Developments in Statistics*, pages 133–146. J.R. Barra, F. Brodeau, G. Romier and B. Van Cutsem, Amsterdam, north holland publishing company edition, 1977. (Cited page 6)

J. De Leeuw and W. J. Heiser. Convergence of correction matrix algorithms for multidimensional scaling. *Geometric representations of relational data*, pages 735–752, 1977. (Cited page 6)

J. De Leeuw and W. J. Heiser. Multidimensional scaling with restrictions on the configuration. *Multivariate analysis*, 5:501–522, 1980. (Cited page 6)

J. De Leeuw and P. Mair. Multidimensional scaling using majorization: SMA-COF in R. *Journal of Statistical Software*, 31(3):1–30, 2009. (Cited page 7)

C. Deutsch and T. Tran. FLUVSIM: a program for object-based stochastic modeling of fluvial depositional systems. *Computers & Geosciences*, 28(4): 525–535, 2002. doi: 10.1016/S0098-3004(01)00075-9. (Cited page 1)

C. V. Deutsch. Fortran programs for calculating connectivity of three-dimensional numerical models and for ranking multiple realizations. *Computers & Geosciences*, 24(1):69–76, 1998. doi: 10.1016/S0098-3004(97)00085-X. (Cited pages 2, 3, 14, 15, and 23)

C. V. Deutsch and T. A. Hewett. Challenges in reservoir forecasting. *Mathematical Geology*, 28(7):829–842, 1996. doi: 10.1007/BF02066003. (Cited page 1)

C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical Software Library and User’s Guide*. Oxford University Press, New York, 1992. ISBN 978-0-19-510015-0. (Cited pages 1 and 20)

A. Fiori and I. Jankovic. On Preferential Flow, Channeling and Connectivity in Heterogeneous Porous Formations. *Mathematical Geosciences*, 44(2): 133–145, Feb. 2012. doi: 10.1007/s11004-011-9365-2. (Cited page 20)

M. Gainski, A. G. MacGregor, P. J. Freeman, and H. F. Nieuwland. Turbidite reservoir compartmentalization and well targeting with 4D seismic and production data: Schiehallion Field, UK. *Geological Society, London, Special Publications*, 347(1):89–102, 2010. doi: 10.1144/SP347.7. (Cited page 19)

A. Galli, H. Beucher, G. L. Loc’h, B. Doligez, and H. Group. The Pros and Cons of the Truncated Gaussian Method. In M. Armstrong and P. A. Dowd, editors, *Geostatistical Simulations*, number 7 in Quantitative Geology and Geostatistics, pages 217–233. Springer Netherlands, 1994. ISBN 978-90-481-4372-6 978-94-015-8267-4. doi: 10.1007/978-94-015-8267-4_18. (Cited page 1)

J. Gómez-Hernández and X.-H. Wen. To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. *Advances in Water Resources*, 21(1): 47–61, 1998. doi: 10.1016/S0309-1708(96)00031-0. (Cited page 1)

J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4):325–338, 1966. doi: 10.1093/biomet/53.3-4.325. (Cited page 6)

F. B. Guardiano and R. M. Srivastava. Multivariate Geostatistics: Beyond Bivariate Moments. In A. Soares, editor, *Geostatistics Tróia ’92*,

- number 5 in *Quantitative Geology and Geostatistics*, pages 133–144. Springer Netherlands, 1993. ISBN 978-0-7923-2157-6 978-94-011-1739-5. (Cited pages 1 and 2)
- L. A. S. Hansen, R. H. T. Callow, I. A. Kane, F. Gamberi, M. Rovere, B. T. Cronin, and B. C. Kneller. Genesis and character of thin-bedded turbidites associated with submarine channels. *Marine and Petroleum Geology*, 67:852–879, Nov. 2015. doi: 10.1016/j.marpetgeo.2015.06.007. (Cited page 19)
- S. M. Hubbard, D. G. Smith, H. Nielsen, D. A. Leckie, M. Fustic, R. J. Spencer, and L. Bloom. Seismic geomorphology and sedimentology of a tidally influenced river deposit, Lower Cretaceous Athabasca oil sands, Alberta, Canada. *AAPG bulletin*, 95(7):1123–1145, 2011. (Cited page 19)
- B. Issautier, S. Fillacier, Y. Le Gallo, P. Audigane, C. Chiaberge, and S. Viseur. Modelling of CO₂ Injection in Fluvial Sedimentary Heterogeneous Reservoirs to Assess the Impact of Geological Heterogeneities on CO₂ Storage Capacity and Performance. *Energy Procedia*, 37:5181–5190, 2013. doi: 10.1016/j.egypro.2013.06.434. (Cited page 19)
- A. K. Jain. *Fundamentals of digital image processing*, volume 3. Prentice-Hall Englewood Cliffs, 1989. ISBN 978-0-13-336165-0. (Cited page 3)
- A. Journel and F. Alabert. New method for reservoir mapping. *Journal of Petroleum technology*, 42(2):212–218, 1990. doi: 10.2118/18324-PA. (Cited page 1)
- A. G. Journel. Beyond covariance: the advent of multiple-point geostatistics. *Geostatistics Banff 2004*, pages 225–233, 2004. (Cited page 2)
- M. King and M. Mark. Flow Simulation of Geologic Models. *SPE Reservoir Evaluation & Engineering*, 2(4):351–367, 1999. doi: 10.2118/57469-PA. (Cited page 1)
- R. Labourdette, J. Poncet, J. Seguin, F. Temple, J. Hegre, and A. Irving. Three-dimensional modelling of stacked turbidite channels in West Africa: impact on dynamic reservoir simulations. *Petroleum Geoscience*, 12(4):335–345, 2006. doi: 10.1144/1354-079306-705. (Cited page 19)
- T. C. Lee, R. L. Kashyap, and C. N. Chu. Building Skeleton Models via 3-D Medial Surface Axis Thinning Algorithms. *CVGIP: Graphical Models and Image Processing*, 56(6):462–478, 1994. doi: 10.1006/cgip.1994.1042. (Cited page 18)
- J. Lin. Divergence measures based on the Shannon entropy. *Information Theory, IEEE Transactions on*, 37(1):145–151, 1991. doi: 10.1109/18.61115. (Cited page 5)
- S. Lopez. *Modélisation de réservoirs chenalisés méandriques : une approche génétique et stochastique*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, 2003. (Cited page 1)
- G. Mariethoz. *Geological stochastic imaging for aquifer characterization*. PhD thesis, Université de Neuchâtel, 2009. (Cited page 4)
- G. Mariethoz, P. Renard, and J. Straubhaar. The Direct Sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, 46(11):W11536, 2010. doi: 10.1029/2008WR007621. (Cited pages 2 and 7)
- E. Meerschman, G. Pirot, G. Mariethoz, J. Straubhaar, M. V. Meirvenne, and P. Renard. A Practical Guide to Performing Multiple-Point Statistical Simulations with the Direct Sampling Algorithm. *Computers & Geosciences*, 52: 307–324, 2012. doi: 10.1016/j.cageo.2012.09.019. (Cited page 2)
- F. Oriani and P. Renard. Binary upscaling on complex heterogeneities: The role of geometry and connectivity. *Advances in Water Resources*, 64:47–61, 2014. doi: 10.1016/j.advwatres.2013.12.003. (Cited page 4)
- Paradigm. SKUA® Software Suite by Paradigm®. <http://www.pdgm.com/products/skua-gocad/>, 2015. (Cited page 7)
- C.-H. Park, C. Beyer, S. Bauer, and O. Kolditz. A study of preferential flow in heterogeneous media using random walk particle tracking. *Geosciences Journal*, 12(3):285–297, Oct. 2008. doi: 10.1007/s12303-008-0029-2. (Cited page 20)
- M. J. Pranter, A. I. Ellison, R. D. Cole, and P. E. Patterson. Analysis and modeling of intermediate-scale reservoir heterogeneity based on a fluvial point-bar outcrop analog, Williams Fork Formation, Piceance Basin, Colorado. *AAPG Bulletin*, 91(7):1025–1051, Jan. 2007. doi: 10.1306/02010706102. (Cited page 19)
- M. Pyrcz, J. Boisvert, and C. Deutsch. ALLUVSIM: A program for event-based stochastic modeling of fluvial depositional systems. *Computers & Geosciences*, 35(8):1671–1685, 2009. doi: 10.1016/j.cageo.2008.09.012. (Cited page 1)
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0. (Cited page 7)
- C. R. Rao. Differential metrics in probability spaces. *Differential geometry in statistical inference*, 10:217–240, 1987. doi: 10.1214/lnms/1215467062. (Cited page 5)
- P. Renard and D. Allard. Connectivity metrics for subsurface flow and transport. *Advances in Water Resources*, 51:168–196, 2013. doi: 10.1016/j.advwatres.2011.12.001. (Cited pages 2, 4, 7, and 23)
- C. Scheidt and J. Caers. Representing Spatial Uncertainty Using Distances and Kernels. *Mathematical Geosciences*, 41(4):397–419, 2009. doi: 10.1007/s11004-008-9186-0. (Cited pages 2 and 6)
- Schlumberger. Petrel E&P Software Platform. <http://www.software.slb.com/products/platform/Pages/petrel.aspx>, 2015. (Cited page 7)
- J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA, 1983. ISBN 978-0-12-637240-3. (Cited page 3)
- R. N. Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27(2):125–140, 1962a. doi: 10.1007/BF02289630. (Cited page 2)
- R. N. Shepard. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, 27(3):219–246, 1962b. doi: 10.1007/BF02289621. (Cited page 2)
- A. Shtuka, P. Samson, and J. L. Mallet. Petrophysical simulation within an object-based reservoir model. In *European 3-D reservoir modelling conference*, pages 47–56, 1996. (Cited page 2)
- J. Straubhaar. MPDS technical reference guide. Technical report, Centre d’hydrogéologie et géothermie, University of Neuchâtel, Neuchâtel, 2011. (Cited page 7)
- J. Straubhaar, P. Renard, G. Mariethoz, R. Froidevaux, and O. Besson. An Improved Parallel Multiple-point Algorithm Using a List Approach. *Mathematical Geosciences*, 43(3):305–328, Mar. 2011. doi: 10.1007/s11004-011-9328-7. (Cited page 7)
- J. Straubhaar, A. Walgenwitz, and P. Renard. Parallel Multiple-Point Statistics Algorithm Based on List and Tree Structures. *Mathematical Geosciences*, 45: 131–147, 2013. doi: 10.1007/s11004-012-9437-y. (Cited page 7)
- S. Strebelle. Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics. *Mathematical Geology*, 34(1):1–21, 2002. doi: 10.1023/A:1014009426274. (Cited page 2)
- P. Tahmasebi, A. Hezarkhani, and M. Sahimi. Multiple-point geostatistical modeling based on the cross-correlation functions. *Computational Geosciences*, 16(3):779–797, 2012. doi: 10.1007/s10596-012-9287-1. (Cited page 2)
- X. Tan, P. Tahmasebi, and J. Caers. Comparing Training-Image Based Algorithms Using an Analysis of Distance. *Mathematical Geosciences*, 46(2):149–169, 2014. doi: 10.1007/s11004-013-9482-1. (Cited pages 2, 6, 13, and 15)
- W. S. Torgerson. Multidimensional scaling: I. Theory and method. *Psychometrika*, 17(4):401–419, 1952. doi: 10.1007/BF02288916. (Cited pages 2 and 6)
- W. S. Torgerson. *Theory and methods of scaling*. John Wiley & Sons, New York, 1958. ISBN 978-0-471-87945-9. (Cited page 6)
- T. T. Tran. Improving variogram reproduction on dense simulation grids. *Computers & Geosciences*, 20(7):1161–1168, 1994. doi: 10.1016/0098-3004(94)90069-8. (Cited page 13)
- S. Viseur. *Simulation stochastique basée-objet de chenaux*. PhD thesis, Institut National Polytechnique de Lorraine, Vandoeuvre-lès-Nancy, France, 2001. (Cited page 1)
- H. Wadell. Volume, shape, and roundness of quartz particles. *The Journal of Geology*, pages 250–280, 1935. (Cited page 23)
- A. W. Western, G. Blöschl, and R. B. Grayson. Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research*, 37(1):83–97, Jan. 2001. doi: 10.1029/2000WR900241. (Cited page 4)
- H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer, 2009. ISBN 978-0-387-98141-3. (Cited page 7)
- D. R. Wilson and T. R. Martinez. Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research*, 6:1–34, 1997. doi: 10.1613/jair.346. (Cited page 5)
- Y. Yin, S. Wu, C. Zhang, S. Li, and T. Yin. A reservoir skeleton-based multiple-point geostatistics method. *Science in China Series D: Earth Sciences*, 52(1): 171–178, 2009. doi: 10.1007/s11430-009-5004-x. (Cited page 2)